# Forecast Optimality Tests in the Presence of Instabilities

Barbara Rossi* and Tatevik Sekhposyan†

October 4, 2011

**Abstract**

This paper proposes forecast optimality tests that can be used in unstable environments. They include tests for forecast unbiasedness, efficiency, encompassing, serial uncorrelation, and, in general, regression-based tests of forecasting ability. The proposed tests are applied to evaluate the rationality of the Federal Reserve Greenbook forecasts as well as a variety of survey-based private forecasts. In addition, we consider whether Money Market Services forecasts are rational. Our robust tests suggest more empirical evidence against forecast rationality than previously found but confirm that the Federal Reserve has additional information about current and future states of the economy relative to market participants.

**Acknowledgements:** We are grateful to Mike McCracken, Andrew Patton, Clara Vega and session participants at the 2011 Joint Statistical Meetings for comments. The views expressed in this paper are those of the authors. No responsibility should be attributed to the Bank of Canada.

**Keywords:** Forecasting, forecast optimality, regression-based tests of forecasting ability, Greenbook forecasts, survey forecasts, real-time data

**J.E.L. Codes:** C22, C52, C53

*Department of Economics, Duke University, 213 Social Sciences, P.O. Box 90097, Durham, NC 27708, USA. Tel.: +1-919-660-1801; fax: +1-919-684-8974; e-mail: brossi@econ.duke.edu

†International Economic Analysis Department, Bank of Canada, 234 Wellington Street, Ottawa, ON, K1A 0G9, Canada. Tel.: +1-613-782-7239; fax: +1-613-782-7658; e-mail: tsekhposyan@bankofcanada.ca

# 1 Introduction

Forecasting is a fundamental tool in economics, as well as statistics, business and other sciences. Judging whether forecasts are good is therefore of great importance, especially since forecasts are used everyday to guide policymakers' and practitioners' decisions. A large literature has provided important insights on how to test whether forecasts are optimal. For example, the seminal works of Granger and Newbold (1986) and Diebold and Lopez (1996) have shown that, under covariance stationarity and a mean square error loss, forecast errors satisfy the following properties: they are mean zero (conditionally and unconditionally), the h-steps-ahead forecast error has zero serial correlation after (h-1) lags, and the unconditional variance of the forecast error is a non-decreasing function of the forecast horizon. If a forecast is such that its forecast errors satisfy such properties, it is deemed optimal.[1] If a forecast does not satisfy these properties, researchers conclude that the model underlying such forecast can be improved.

However, one of the fundamental assumptions tacitly underlying the existing literature is that of covariance stationarity. Only very recently researchers have become concerned about the consequences of the stationarity assumption in performing inference regarding predictive ability.[2] Giacomini and Rossi (2010) have developed methods to perform inference on forecast comparisons when the forecasting ability may be affected by instabilities. Besides the forecasting ability, another important issue that forecasters face in practice is to determine whether forecasts are rational or optimal, and that might also be affected by instabilities. In fact, several studies evaluate the robustness of forecast rationality in sub-samples (e.g. Croushore 1998, Patton and Timmermann, 2011, Croushore, 2011). However, while in some cases the choice of the sub-samples may be guided by economic considerations (e.g. sub-samples associated with structural breaks identified by the Great Moderation), in many cases the choice of sub-samples may be ad-hoc. Even when the choice is guided by economic considerations, it may be important to assess the robustness of the empirical results to other sub-samples, as there might be multiple breaks in the data, or the date of the break might be completely unknown or relatively uncertain.

This paper proposes forecast optimality tests that are robust to the presence of instabilities. We consider a framework where forecasts are produced either with a recursive, fixed, or a rolling estimation scheme, and the size of the forecast window is large relative to the

[1] Covariance stationarity is a key assumption for the last property: it would not need to be true otherwise.
[2] See the discussion in Rossi (2011b).

sample size. We propose a "Fluctuation Optimality" test, which is based on estimating and testing for forecast optimality in rolling windows over the out-of-sample forecast portion of the data. By using rolling windows we avoid averaging out instabilities, and our tests have greater power to reject forecast optimality than traditional tests when rationality is present only in a sub-sample of the data. Our "Fluctuation Optimality" test can be applied to study forecast unbiasedness, rationality, encompassing, as well as serial uncorrelation, among other regression-based tests of forecasting ability.

This paper is closely related to Giacomini and Rossi (2010) and West and McCracken (1998). Giacomini and Rossi (2010) have proposed a "Fluctuation test" to compare forecasting models in the presence of instabilities. While our "Fluctuation Optimality" test is inspired by their work, there are several differences between their framework and ours. Their framework compares models' *relative* forecasting performance and is focused primarily on a rolling window estimation where the size of the window is fixed (i.e. finite). We are instead interested in measures of *absolute* predictive ability and tests for forecast optimality. Our framework focuses on an estimation window size that is a fixed fraction of the total sample size, which allows us to take into account parameter estimation error. The latter framework is similar to that of West and McCracken (1998). The difference between our tests and the ones proposed by West and McCracken (1998) is that the latter is based on measures of average forecasting ability in the out-of-sample portion of the data. We show that these tests do not have the power to detect lack of forecast optimality in situations when there are optimality breakdowns over time.

We demonstrate the usefulness of our procedures in two empirical applications. The first focuses on evaluating the rationality of the Federal Reserve's Greenbook forecast of inflation as well as the private sector's forecasts provided by the Survey of Professional Forecasters and Blue Chip Economic Indicators. We revisit the empirical analysis in Romer and Romer (2000), Patton and Timmermann (2011), and Croushore (2011) in a framework that is robust to the presence of instabilities. We then reconsider Romer and Romer's (2000) hypothesis that the Federal Reserve has an information advantage in forecasting inflation beyond what is known to the private forecasters, again using our framework robust to time-variation. The second empirical application focuses on testing forecast rationality in the Money Market Services survey of economic indicators. Using traditional Mincer and Zarnowitz's (1969) regressions, several papers in the literature have concluded that the Money Market Services forecasts are rational (e.g. Balduzzi, Elton and Green, 2001, and Urich and Wachtel, 1984). Our results show instead that it is possible to reject forecast rationality of the Money Market

Services forecasts at some point in time.

The paper is structured as follows. The first section presents the econometric methodology. The second section studies the size and power of our "Fluctuation Optimality" test in small samples. The third section discusses the empirical applications, and the last section concludes.

# 2 The Econometric Framework

The main objective of this paper is to test whether the $h-$steps ahead, out-of-sample direct forecasts for the variable $y_t$, which we assume for simplicity to be a scalar, are optimal. We assume that the forecasts are based on a model that is characterized by the $(k \times 1)$ parameter vector $\gamma$. The forecasts are obtained by dividing the sample of size $T+h$ observations into an in-sample portion of size $R$ and an out-of-sample portion of size $P$. The sequence of $P$ out-of-sample forecast errors depend on the realizations of the forecasted variable and on the in-sample parameter estimates, $\widehat{\gamma}_{t,R}$. According to usual forecasting practices, we assume that these parameters are estimated in either one of three ways: (i) only once, using a sample including data indexed $1,...,R$ (fixed scheme); (ii) re-estimated at each $t=R,...,T$ over a window of $R$ observations including data indexed $t-R+1,...,t$ (rolling scheme); (iii) re-estimated at each $t=R,...,T$ over a window of $t$ observations including data indexed $1,...,t$ (recursive scheme).

Let the forecast error associated to the h-steps ahead forecast made at time $t$ be denoted by $v_{t+h}$. For example, in the case of a simple linear regression model with h-periods lagged regressors $x_t$, the forecast error associated with the direct forecast is: $v_{t+h} = y_{t+h} - \widehat{\gamma}_{t,R} x_t$.

We focus on testing for forecast optimality in the framework developed by West and McCracken (1998). Consider the general regression:

$$v_{t+h} = g_t' \cdot \theta + \eta_{t+h}, \ t = R, ..., T, \tag{1}$$

where $\theta$ is a $(q \times 1)$ parameter vector, $v_{t+h}$ is the realized forecast error, and $g_t$ is a $(q \times 1)$ vector of variables known at time $t$ such that $E\left(g_t g_t'\right)$ is full rank. West and McCracken's (1998) forecast rationality tests focus on testing the null hypothesis:

$$H_0 : \theta = \theta_0 \text{ vs. } H_A : \theta \neq \theta_0.$$

Let $\widehat{\theta}_P$ denote the estimate of $\theta$ in regression (1). Consider the following Wald test:

$$\mathcal{W}_P = \left(\widehat{\theta}_P - \theta_0\right)' \ \widehat{V}_{\theta,P}^{-1} \left(\widehat{\theta}_P - \theta_0\right), \tag{2}$$

where $\widehat{V}_{\theta,P}$ is a consistent estimate of the long run variance of $\widehat{\theta}_P$. West and McCracken (1998) show that in order to estimate the long run variance consistently it is important to correct the estimate of the variance by parameter estimation error as well as, in some cases, by adding an additional explanatory variable to equation (1).

West and McCracken (1998) consider the following leading cases:

(i) forecast unbiasedness tests, where $g_t = 1$;

(ii) forecast efficiency, where $g_t = y_{t+h|t}$. A special case is the forecast rationality test in Mincer and Zarnowitz (1969), where $g_t = [1 \ y_{t+h|t}]$, $\theta = [\alpha, \beta]'$, and typically a researcher is interested in testing whether $\alpha$ and $\beta$ are jointly zero;

(iii) forecast encompassing tests, where $g_t$ is the forecast of the encompassed model;

(iv) serial uncorrelation tests, where $g_t = v_t$.

We refer to all these tests under the maintained assumption that $\theta_0 = 0$ as "tests for forecast optimality"; the zero restriction on the parameter under the null hypothesis ensures that the forecast errors are truly unpredictable given the information set available at the time when the forecast is made.

Our main interest is testing forecast optimality in the presence of instabilities. In fact, in the presence of instabilities, tests that focus on the average out-of-sample performance of a model may be misleading, as they may average out instabilities. Instead, we consider the following rolling regression. Let $\widehat{\theta}_{t,R}$ be the parameter estimate in regression (1) computed over rolling windows of size $m$ (without loss of generality, we assume $m$ to be an even number). That is, consider estimating regression (1) using data from $t - m + 1$ up to $t$, for $t = m, ..., P$. Also, let the Wald test in the corresponding regressions be defined as:

$$\mathcal{W}_{t,m} = \left(\widehat{\theta}_t - \theta_0\right)' \ \widehat{V}_{\theta,t}^{-1} \left(\widehat{\theta}_t - \theta_0\right), \text{ for } t = m, ..., P, \tag{3}$$

where $\widehat{V}_{\theta,t}$ is a HAC estimator of the asymptotic variance of the parameter estimates in the rolling windows obtained as in West and McCracken (1998). We refer to $\mathcal{W}_{t,m}$ as the "Fluctuation Optimality" test.[3]

Let $\mathcal{L}_{t+h} \equiv v_{t+h} g_t$. We make the same assumptions as in West and McCracken (1998). Note that while the assumptions are the same, our results are stronger since the limiting behavior of the re-scaled test statistics that we propose obeys a Functional Central Limit Theorem.

[3] In the construction of the test we associate the end of period date of the fixed window $m$ with the parameter estimate $\theta_t$. In fact, that need not necessarily be the case. If one prefers, one can choose to associate the mid-period date of the fixed window $m$ with the parameter estimate, for example.

Let the forecasts be based on the $(k \times 1)$ true parameter vector $\gamma^*$. Let $\mathcal{L}_{t+h,\gamma} \equiv \partial \mathcal{L}_{t+h}(\theta_0)/\partial \gamma$, $v_{t,\gamma} \equiv \partial v_t(\theta_0)/\partial \gamma$, $g_{t,\gamma} \equiv \partial g_t(\theta_0)/\partial \gamma$, $\mathcal{L}_\gamma \equiv E(\mathcal{L}_{t+h,\gamma})$, $||.||$ denote the Euclidean norm.

*Assumption 1:*

*(1) In some neighborhood N around $\gamma^*$, and with probability 1, $v_t(\gamma)$ and $g_t(\gamma)$ are measurable and twice continuously differentiable; (b) $E(v_{t+h}g_t) = 0$; (c) $E(v_t v_{t,\gamma}) = 0$; (d) $E(v_{t+h}g_{t,\gamma}) = 0$; (e) $E(g_t g_t')$ is of rank l.*

*(2) The estimate $\widehat{\gamma}_t$ satisfies $\widehat{\gamma}_t - \gamma^* = B_t H_t$ where $B_t$ is $(k \times q)$ matrix and $H_t$ is $(q \times 1)$ with (a) $B_t \underset{p}{\rightarrow} B$ with rank k; (b) $H_t = t^{-1}\sum_{s=1}^{t} h_s(\gamma^*)$ for the recursive estimation method or $H_t = R^{-1}\sum_{s=t-R+1}^{t} h_s(\gamma^*)$ for the rolling, and $H_t = R^{-1}\sum_{s=1}^{R} h_s(\gamma^*)$ for the fixed estimation methods for a $(q \times 1)$ orthogonality condition $h_s(\gamma^*)$; (c) $E(h_s(\gamma^*)) = 0$; (d) $h_t$ is measurable and continuously differentiable in the neighborhood N defined in (1).*

*(3) There is a finite constant D such that for all t $sup_{\gamma \in N}|\partial^2 v_t(\theta_0)/\partial\gamma\partial\gamma'| < m_t$ for a measurable $m_t$ such that $E(m_t^4) < D$ and the same holds when $v_t$ is replaced by any element of $g_t$.*

*(4) Let $w_t \equiv [v_{t,\gamma}', vec(g_{t,\gamma})', v_t', g_t', h_t']'$. For some $d > 1$, $sup_t E||w_t||^{8d} < \infty$; (b) $w_t$ is strong mixing with mixing coefficient of size $-3d/(d-1)$; (c) $w_t$ is fourth order stationary; (d) $S_{ff} \equiv \Sigma_{j=-\infty}^{\infty}\Gamma_{ff}(j)$ is positive definite, where $\Gamma_{ff}(j) = E(\mathcal{L}_t \mathcal{L}_{t-j})$.*

*(5) $lim_{T\rightarrow\infty} m/P = \mu \in (0,\infty)$ as $m \rightarrow \infty, P \rightarrow \infty$, $h < \infty$, and $lim_{T\rightarrow\infty} P/R = \pi \in [0,\infty)$ as $R, P \rightarrow \infty$.*

The first assumption requires several moment conditions to hold. The second assumption allows the in-sample parameter estimates to be obtained by general estimation procedures such as Ordinary Least Squares (OLS), maximum likelihood, and GMM, for example. The third assumption imposes boundedness conditions. The fourth assumption allows the data to be mixing and heterogenous with positive asymptotic variances. Finally, the last assumption requires $P, R$ and $m$ to be large relative to the sample size (in particular, relative to the horizon, $h$) and ensures the consistency of the out-of-sample test statistics. The assumption on $\pi$ accommodates fixed, rolling, and recursive estimation schemes.

**Theorem 1 (Main Proposition)** *Under Assumption 1,*

$$\mathcal{W}_{t,m} \Rightarrow \mu^{-1}\left[\mathcal{B}_q(\tau+\mu/2) - \mathcal{B}_q(\tau-\mu/2)\right]'\left[\mathcal{B}_q(\tau+\mu/2) - \mathcal{B}_q(\tau-\mu/2)\right],$$

*where $t = [\tau P]$, $m = [\mu P]$ and $\mathcal{B}_q(\cdot)$ is a standard q-dimensional Brownian motion. We* *reject the null hypothesis:*

$$H_0 : E\left(\widehat{\theta}_t\right) = \theta_0 \text{ for all } t = m, ..., P \tag{4}$$

*if* $max_t\ \mathcal{W}_{t,m} > \kappa_{\alpha,q}$*, where* $\kappa_{\alpha,q}$ *are the critical values at the* $100\alpha\%$ *significance level and are reported in Table 1 for various values of* $\mu = [m/P]$ *and number of restrictions, q.*

**Proof of Theorem 1.** It suffices to verify that Assumptions C1, C2 and A5 in Corollary 3.1 of Wooldridge and White (1988) are satisfied. Assumptions C1 and A5 follow directly from West and McCracken (1998). To show that assumption C2 holds, note from West and McCracken (1998) that $\Sigma_{t=R}^{T} H_t$ is a weighted average of $h_t$ with bounded coefficients. ■

INSERT TABLE 1 HERE

We obtain the critical values by Monte Carlo simulation. The critical values at significance level $100\alpha\%$ are such that

$$\Pr\left\{\sup_{\tau}\left[\mathcal{B}_q\left(\tau+\mu/2\right)-\mathcal{B}_q\left(\tau-\mu/2\right)\right]'\left[\mathcal{B}_q\left(\tau+\mu/2\right)-\mathcal{B}_q\left(\tau-\mu/2\right)\right]\mu^{-1} > \kappa_{\alpha,q}\right\} = \alpha. \tag{5}$$

There are two special cases of our methodology. First, we should note that the testing procedure reduces to a similar procedure as in Giacomini and Rossi (2010) in the special case where $q = 1$ and the estimation window size is finite. In practice, this requires attention in the construction of the asymptotic variance: provided the West and McCracken's (1998) unadjusted asymptotic variance is used to construct the test that we propose, a t-statistic version of our test statistic has exactly the same distribution as Giacomini and Rossi (2010), even though in the latter the window size is fixed relative to the total sample size. That is, in the special case of a two-sided t-ratio test on the s-th parameter, $\theta_0^{(s)}$, the test can be obtained as:

$$t_{t,m} \equiv \left(\widehat{\theta}_t^{(s)} - \theta_0^{(s)}\right)\widehat{V}_{\theta^{(s)},t}^{-1/2}, \text{ for } t = m, ..., P,$$

where $\widehat{V}_{\theta^{(s)},t}$ is element in the s-th row and s-th column of $\widehat{V}_{\theta,t}$. We reject the null hypothesis $H_0 : E\left(\widehat{\theta}_t^{(s)}\right) = \theta_0^{(s)}$ for all $t = m, ..., P$ at the $100\alpha\%$ significance level if $\max_t \left|\left(\widehat{\theta}_t^{(s)} - \theta_0^{(s)}\right)\widehat{V}_{\theta^{(s)},t}^{-1/2}\right| > k_\alpha$, where $k_\alpha$ are the two-sided test critical values provided by Giacomini and Rossi (2010).

Second, in the case of a linear regression model equation (2) is the same as that considered in West and McCracken (1998). In order to correctly take into account parameter estimation error, we recommend to either adjust the variance explicitly or consider the auxiliary regressions suggested by West and McCracken (1998). The explicit variance adjustment pertains to the cases of mean prediction error and efficiency tests described earlier. In these cases the variance is multiplied by an adjustment factor $\lambda$, which is: (i) $\lambda = 1 + \pi$ for the fixed estimation scheme; (ii) $\lambda = 1 - \pi^2/3$ when $\pi \leq 1$ and $\lambda = 2/(3\pi)$ when $\pi > 1$ for the rolling estimation scheme; (iii) $\lambda = 1$ for the recursive estimation scheme. In the cases of the encompassing and serial uncorrelation tests, we consider alternative regressions similar to that in equation (1) with the following adjustment: consider $v_{t+h} = \tilde{g}_t' \cdot \theta + \eta_{t+h}$, where $\tilde{g}_t = (g_t, W_t)$ and $W_t$ denotes the set of regressors that enter the predictive regression of the serial uncorrelation test and that of the particular regressors common to the encompassing (as opposed to the encompassed) models respectively. Note that this auxiliary regression is necessary only when the forecasts are obtained based on fixed and rolling estimation schemes, or under the recursive scheme with conditionally heteroskedastic forecast errors $v_{t+h}$.

It should be noted that, as shown in Rossi (2011a), our framework can encompass the hypothesis of forecast optimality for more general definitions of optimality. As Patton and Timmermann (2010) suggest, under certain regularity conditions forecast optimality tests can reduce to those considered in this paper for arbitrary loss functions (symmetric or asymmetric) if one adheres to the definition of "generalized forecast error" and/or changes the probability measure.

# 3 Monte Carlo Analysis

We study the small sample performance of the methods that we propose in a series of Monte Carlo experiments inspired by West and McCracken (1998). Let the Data Generating Process (DGP) be: $y_t = \rho y_{t-1} + \varepsilon_t$, where $\rho = 0.5$, $\varepsilon_t \sim iid(0,1)$, and $y_0$ is drawn from its unconditional distribution, a normal with zero mean and variance $(1-\rho^2)^{-1}$. In each sample, $t = 1, ..., T$, we split the data into $T = P + R + 1$, and we utilize either a rolling or a recursive scheme to generate $P$ one-step ahead, out-of-sample forecasts, $y^f_{R+1}, ..., y^f_T$, and forecast errors: $\varepsilon^f_{R+1}, ..., \varepsilon^f_T$, where $\varepsilon^f_{t+1} \equiv y_{t+1} - y^f_{t+1}$, for $t = R, ..., T$. The advantage of using the same DGP as West and McCracken (1998) is that we can directly compare our results to theirs.

First, we consider the size properties of our tests. We focus on the same four tests in West and McCracken (1998). Consider the following regression model:

$$X_t = \alpha + \beta Z_t + \gamma W_t + U_t \tag{6}$$

The four tests we consider can be described as follows:

(i) tests for zero mean prediction error (or forecast unbiasedness). They test whether the mean of the sequence of forecast errors is zero and are implemented by a two-sided t-test on whether $\alpha = 0$ in regression (6), where $X_t = \varepsilon_t^f$ and there are no regressors other than the constant. Let $\widehat{\alpha} = P^{-1} \sum_{t=R}^{T} \varepsilon_{t+1}^f$ and $\widehat{\sigma}^2 = P^{-1} \sum_{t=R}^{T} \left(\varepsilon_{t+1}^f\right)^2$. We consider two type of tests: an uncorrected t-test, $t_\alpha = P^{1/2}\widehat{\alpha}\widehat{\sigma}^{-1}$, and a t-test that utilizes West and McCracken's (1998) correction: $t_\alpha^c = P^{1/2}\widehat{\alpha}\widehat{\sigma}^{-1}\lambda^{-1/2}$, where $\lambda = 1$ for the recursive scheme and, for the rolling scheme, $\lambda = 1 - \pi^2/3$ when $\pi \leq 1$ and $\lambda = 2/(3\pi)$ when $\pi > 1$.

(ii) tests for forecast efficiency, which are implemented by a two-sided t-test on whether $\beta = 0$ in the regression (6), where $X_t = \varepsilon_t^f$, $Z_t = y_t^f$, and no other regressor is included. Let $\widehat{\beta}$ be the OLS coefficient estimate in the latter regression, and $\widehat{\sigma}_\beta^2$ be its estimated standard error. We consider two type of tests: an uncorrected t-test, $t_\beta = \widehat{\beta}\widehat{\sigma}_\beta^{-1}$, and a t-test that utilizes West and McCracken's (1998) correction: $t_\beta^c = \widehat{\beta}\widehat{\sigma}_\beta^{-1}\lambda^{-1/2}$, for the same values of $\lambda$ as in (i).

(iii) tests for forecast encompassing. Let the encompassing model be a regression of $y_t$ onto $y_{t-1}$ and $y_{t-2}$, whose forecast errors are denoted by $\widetilde{\varepsilon}_{R+1}^f, ..., \widetilde{\varepsilon}_T^f$. Let the encompassed model be a regression of $y_t$ onto $y_{t-2}$, whose corresponding forecasts are denoted respectively by $\widehat{y}_{R+1}^f, ..., \widehat{y}_T^f$. The forecast encompassing test is typically implemented by a two-sided t-test on whether $\beta = 0$ in the regression (6) where: $X_t = \widetilde{\varepsilon}_t^f$ and $Z_t = \widehat{y}_t^f$. The West and McCracken's (1998) corrected version of the test involves a two-sided t-test on whether $\beta = 0$ in the regression: $X_t = \widetilde{\varepsilon}_t^f$, $Z_t = \widehat{y}_t^f$, and $W_t = y_{t-2}$.

(iv) tests for serial uncorrelation of the forecast errors, which are implemented by a two-sided t-test on whether $\beta = 0$ in the regression (6), where: $X_t = \varepsilon_t^f$, $Z_t = \varepsilon_{t-1}^f$. The West and McCracken's (1998) corrected version of the test involves a two-sided t-test on whether $\beta = 0$ in the regression (6), where: $X_t = \varepsilon_t^f$, $Z_t = \varepsilon_{t-1}^f$ $W_t = y_{t-1}$.

In addition, we consider our proposed "Fluctuation Optimality" test, equation (3), implemented in rolling regressions over the out-of-sample period with a rolling window size equal $m = 50$. We implement our test with and without the corrections suggested by West and McCracken (1998) as well, as described above. The number of Monte Carlo replications is 5,000.

Tables 2 and 3 report results for the recursive and the rolling estimation schemes, respectively. Panel A reports results for testing forecast unbiasedness, panel B for forecast efficiency, panel C for forecast encompassing and panel D for serial uncorrelation. The tables shows that the empirical rejection frequencies of our proposed tests (reported in the column labeled “Fluctuation Test”) as well as those of the traditional tests (reported in the column labeled “Traditional Test") are close to the nominal value except in very small sample sizes. The size distortions in small samples are mild for the recursive scheme for both tests, and more substantial for the rolling scheme. We conclude that researchers facing sample sizes smaller than 100 observations should rely on the recursive estimation scheme. For samples larger than 100 observations both tests perform quite well for both estimation schemes.

INSERT TABLES 2 AND 3 HERE

In order to evaluate the power of our test in the presence of time variation, we consider experiments based on three DGPs. DGP 1: $y_t = \rho y_{t-1} + \varepsilon_t + b_t$, where $\rho = 0.5$, $\varepsilon_t \sim iid(0,1)$, $b_t = b \cdot 1(1 < t \leq 345) - b \cdot 1(345 < t < T)$, $b = \{0, 0.1, ..., 1\}$, and predictions are constructed based on the true model. DGP 1 is used to assess the power of the “Fluctuation Optimality" test for the mean prediction error case. To assess the power properties of the efficiency and encompassing tests we consider DGP 2, which is similar to DGP 1 except $b = \{0, 0.5, ..., 5\}$. In this case too, the forecasts are constructed based on the true model. DGP 3 is used for the power experiment for the “Fluctuation Optimality" test for serial uncorrelation. DGP 3 is specified as $y_t = \rho_t y_{t-1} + \varepsilon_t$, where $\rho_t = \rho + b \cdot 1(1 < t \leq 345) - b \cdot 1(345 < t < P)$ and $b = \{0, 0.05, ..., 0.5\}$. In addition, the forecasts and forecast errors are based on a mis-specified model: $y_t = a + \varepsilon_t$. In all cases, $T = 400$, $R = 300$ and $m = 50$.

The power comparisons are reported in Table 4. The table shows that, for the cases that we consider, the traditional tests do not have power to reject the null hypothesis, and in fact their rejection frequencies approach zero under the alternative hypothesis, whereas our proposed tests do have substantial power.

INSERT TABLE 4 HERE

# 4 Private Sector’s versus Federal Reserve’s Forecasts

The quality of private sector’s forecasts relative to the internal forecasts of the Federal Reserve has been frequently considered in the literature. In important contribution, Romer and Romer (2000) showed that the Federal Reserve has more information relative to the private sector when forecasting inflation. Hence, it would be optimal for a third party with access to both forecasts to put all the weight on the forecasts provided by the Federal Reserve and zero weight on the ones provided by the commercial forecasters.

We revisit the existing empirical evidence from two points of view. First, we consider the optimality of private sector's as well as the Federal Reserve's Greenbook forecasts, as in Romer and Romer (2000), Faust and Wright (2010), Patton and Timmermann (2011) and Croushore (2011), among others. These papers have found that forecast rationality tests for the various, competing inflation forecasts are sensitive to the sub-sample period used for forecast evaluation. The novelty of our approach is to study whether forecast rationality holds by using our "Fluctuation Optimality" test robust to instabilities. One of the advantages of our approach is that it does not require researchers to know or impose a sub-sample date a-priori. Second, we evaluate whether Romer and Romer's (2000) finding that Federal Reserve forecasts are superior to private sectors' forecasts continues to hold when we allow for instabilities.

We consider the Federal Reserve's inflation forecasts provided in the Greenbook and compare them with two commercial forecasts: the Blue Chip Economic Indicators (BCEI) and the Survey of Professional Forecasters (SPF). In what follows, we describe the data from each of the sources.

Greenbook forecasts are made by the staff of the Federal Reserve Board of Governors prior to each Federal Open Market Committee (FOMC) meeting. The Greenbook provides quarterly forecasts (from contemporaneous up to nine quarters) for a variety of economic indicators and for several forecast horizons under a maintained assumption about monetary policy; the forecast horizons can vary depending on when the forecasts were made. We consider only forecasts up to five quarters to ensure a sample large enough for inference. We focus on inflation forecasts provided by the Greenbook, which are measured by (annualized) quarter-over-quarter GNP deflator growth rates from 1965 to 1991 and by (annualized) quarter-over-quarter GDP deflator growth rates afterwards. Greenbook forecasts are available only with a five-year lag. Thus, our current sample includes data up to 2005:IV. The data are provided by the Federal Reserve Bank of Philadelphia, which matches the timing of the Greenbook forecasts with that of the SPF. The database includes forecasts from four of the annual FOMC meetings whose the date is closest to the middle of the quarter.[4] In order to make the two data sets comparable, we omit the first 3 years of observations and start the series at 1968:IV.

[4] Greenbook forecasts can be obtained from the Philadelphia Fed web-site at http://www.philadelphiafed.org/research-and-data/real-time-center/greenbook-data/, while the SPF forecasts are provided by the same source at http://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/.

The Survey of Professional Forecasters (SPF) provides forecasts for inflation as well as a variety of economic fundamentals at the quarterly frequency. These include nowcasts (forecasts of the current quarter) as well as forecasts up to four-quarter-ahead. We use the forecasts of (annualized) quarterly GNP/GDP deflator growth rates whose timing is consistent with that of the Greenbook forecasts. The survey is conducted roughly at the end of every second month in the quarter, and it includes 34 professionals' forecasts. The series start at 1968:IV. We use the median forecast and terminate our series at 2005:IV to obtain a data set spanning the same period of time as that of the Greenbook.

The Blue Chip Economic Indicators (BCEI) provides monthly forecasts of quarterly economic series starting from 1980. It is a survey-based forecast database where about 50 U.S. business economists participate each month. Though this is a monthly series, in order to match the Greenbook and SPF forecasts we take only four forecasts per year corresponding to the mid-quarter, i.e. February, May, August, and November, from 1980 to 2005.[5]

To evaluate the Greenbook, SPF and Blue Chip forecasts, we use realized values of (annualized) quarter-over-quarter growth rates of the GNP/GDP deflator constructed from the quarterly vintages in the real-time data set discussed by Croushore and Stark (2001). Our forecast evaluation approach is consistent with that in Romer and Romer (2000), who use the second revision as the benchmark for forecast evaluation, i.e. the specific quarter data available at the last month of the consecutive quarter. Given the real-time nature of the data set, the way we construct the (annualized) quarter-over-quarter GNP/GDP deflator based inflation rate is as follows. For example, the deflator for 1968:IV uses the 1969:I vintage and applies the following transformation: $400ln(PGDP68:IV/PGDP68:III)$. We do so for all the vintages up to 2007:IV, then take the diagonal elements of the resulting matrix. This way we obtain a real-time, annualized measure of the quarter-over-quarter inflation rate of the previous quarter, which we use to evaluate the nowcast or the corresponding $h$-quarter-ahead forecast over time.[6]

Figure 1 compares the Greenbook forecasts with those of the SPF and BCEI. Each panel in Figure 1 corresponds to a forecast horizon, where the horizon $h$ ranges from 0 to 5. $h = 0$ corresponds to the nowcast of inflation. The figure also plots the realized values for inflation at each horizons (reported by the solid line, labeled "actual"). In the figure, not all forecasts

[5] Although the BCEI forecasts are available from August 1976, the forecasts for the initial four years are for annual changes in key economic variables as opposed to quarterly, thus we omit the earlier period.

[6] In the real-time data set provided by the Philadelphia Fed, the observation for 1995:IV is missing in the vintage of 1996:I. We use the value available in the vintage of 1996:II as a substitute value.

have the same starting point. In addition, there are several missing values at several horizons across the different sources, and even for the same source depending on when the forecast has been made. However, overall, the forecasts appear to be correlated with each other: the correlation coefficients between the Greenbook and the private sector's forecasts range from 0.94 to 0.96 across various horizons. Table 5 reports the mean squared forecast errors (MSFE) for the forecast plotted in Figure 1. It appears that the SPF forecasts are inferior to those of the Greenbook at all horizons whereas the BCEI forecasts appear to be superior. However, as we show further, there is substantial evidence of instabilities, and the difference is most likely associated by the sample period that BCEI covers.

INSERT FIGURE 1 AND TABLE 5 HERE

To evaluate the forecast performance, we consider the following regression:

$$\pi_{t+h} = \alpha + \delta\hat{\pi}_{t+h,t} + \epsilon_{t+h}, \tag{7}$$

where $\pi_{t+h}$ is the realized inflation rate, $\hat{\pi}_{t+h,t} = E_t\pi_{t+h}$ is the inflation expectation for $t+h$ based on the information available at the time $t$, $h$ is the forecast horizon and $\epsilon_{t+h}$ is a forecast error. We consider $h = 0, 1, ..., 6$ for the Greenbook forecast, $h = 0, 1, ..., 5$ for the BCEI, and $h = 0, 1, ..., 4$ for the SPF. For the Greenbook, the choice of $h$ is constrained by the need to have a sample size large enough for inference. The choice of $h$ for the BCEI and SPF is dictated by data availability. In order to test forecast optimality in the framework discussed in Section 2, where the test involves zero restrictions on the parameters, we rewrite equation (7) as follows:

$$\pi_{t+h} - \hat{\pi}_{t+h,t} = \alpha + \beta\hat{\pi}_{t+h,t} + \epsilon_{t+h}, \tag{8}$$

where $\beta = \delta - 1$. We present results for both traditional forecast rationality tests as well as the "Fluctuation Rationality" test that we propose. The former relies on the maintained assumption that the parameters of the regression are time invariant and it is implemented using a simple Wald-type test in equation (8), where the parameters are estimated by OLS; we use a HAC variance estimate (Newey and West, 1987) with a bandwidth equal to $P^{1/4}$. Our proposed test instead assesses whether the parameters equal the values implied by optimal forecasts at any given point in time, and it is robust to instabilities. The test is implemented as in eq. (3), where $\widehat{\theta}_t$ are the OLS estimates of $\alpha$ and $\beta$ from equation (8) in rolling regressions with a window size $m = 60$.

The column labeled "Fluctuation" reports the test statistic $\mathcal{W}_{t,m}$ in eq. (3) and the column labeled "Traditional" reports the test statistic $\mathcal{W}_P$ in eq. (2); both are reported for several horizons $h$, listed in the first column. Asterisks denote significance at the 5% significance level. Table 6 suggests that traditional forecast rationality tests fail to reject the null hypothesis of forecast rationality at the 5% significance level for the Greenbook and SPF forecasts, whereas they reject forecast rationality for the BCEI forecasts. However, as we show later, this difference in the results highly depends on the evaluation period, as the sample for the BCEI forecasts starts much later. It is in fact during a period of time when the Greenbook and SPF forecasts fail the rationality test as well. In contrast, the Fluctuation Rationality test rejects the null hypothesis of rationality for all forecasts.

INSERT TABLE 6 HERE

Figures 2-4 plot our test statistic $\mathcal{W}_{t,m}$ in eq. (3) together with the critical values at the 5% significance level. Though the timing on the horizontal axis is suggestive rather than precise, it nevertheless provides some useful information about the timing of the forecast rationality breakdowns. Figure 2 focuses on the Greenbook's forecasts. The figure shows three substantial breakdowns: the first two are associated with the beginning and the end of 1990s. It appears that the forecasts deteriorate over the 1990s and rationality tends to recover by the 2000s. However, for almost all forecast horizons with the exception of five quarters ahead, forecast rationality breaks down again in 2005. Overall, it appears that the empirical evidence in favor of forecast rationality supported by the traditional forecast rationality tests, reported in Table 6, is driven mainly by the good performance of the Greenbook forecasts at the beginning of our sample.

INSERT FIGURES 2, 3 AND 4 HERE

Figure 3 plots the Fluctuation Rationality test for the BCEI forecasts and Figure 4 reports the same test for the SPF forecasts. Figure 4 suggests that the empirical evidence on forecast rationality for SPF forecasts is qualitatively similar to that of the Greenbook. However, the recovery of forecast rationality during the first half of 2000s is less pronounced for SPF than for the Greenbook. By comparing Figure 4 with Figure 3, we note that they behave similarly in the overlapping part of the evaluation period. This suggests that the traditional forecast rationality test results for the BCEI reported in Table 6 are different from the other forecasts solely due to the different sample period. The BCEI forecasts are overall qualitatively similar to the SPF forecasts, with a notable exception: the non-existence of the breakdown of forecast rationality in the BCEI forecasts in 2005.

In general, the empirical evidence in Figures 2-4 does not support forecast rationality for any of the forecasts at any horizons. In addition, if one conditions the forecast optimality results to the recession dates, there appears to be a correlation between the forecast optimality breakdowns prior to the recessions. The recession dates in our out-of-sample evaluation period are: 1990:III-1991:I, 2001:I-2001:IV, 2007:IV-2009:II. If we look at the Greenbook forecasts in Figure 2, as well as the SPF forecasts in Figure 4, we can see that the test statistics indicate a deterioration in forecast optimality at or running up to a recession. Certainly, this is a conjecture, as opposed to formal testing, but the forecast optimality breakdown at the end of the sample (prior to financial crises) is indeed very interesting.

Our second objective is to assess whether the Federal Reserve has an information advantage over private sector's forecasts. To do so, we consider the following regression:

$$\pi_{t+h} - \hat{\pi}^{i}_{t+h,t} = \delta + \beta_g \hat{\pi}^{G}_{t+h,t} + \beta_i \hat{\pi}^{i}_{t+h,t} + \nu_{t+h}, \tag{9}$$

where $\hat{\pi}^{g}_{t+h,t}$ is the Greenbook forecast and $\hat{\pi}^{i}_{t+h,t}$, $i = SPF, BCEI$ denote the SPF and BCEI forecasts, respectively. The Federal Reserve forecasts are useful beyond that of the private sector in predicting inflation if and only if $\beta_g \neq 0$. We test this hypothesis both with the traditional tests as well as with our robust Fluctuation-type test. The latter test is implemented as in eq. (3), where $\widehat{\theta}_t$ are the OLS estimates of $\delta$ and $\beta_g$ from equation (9) in rolling regressions with a window size $m = 60$.

The results are reported in Table 7. The table reports the traditional test statistics (column labeled "Traditional") and the Fluctuation-type test statistic (column labeled "Fluctuation"); asterisks denote significance at the 5% level. According to the table, both the traditional tests and the Fluctuation-type test suggest statistically significant evidence that the Federal Reserve has additional information relative to the private sector's forecasts. Figure 5 sheds additional light on this conclusion. The figure plots the Fluctuation-type test statistics over time and shows that the information advantage of the Federal Reserve has deteriorated after 2003. In fact, the rejections of the hypothesis of no information advantage of the Federal Reserve based on the Fluctuation test appear mostly at the beginning of the sample. The result holds also for both commercial forecasts, that is the BCEI and the SPF.

INSERT TABLE 7 AND FIGURE 5 HERE

Figure 6 plots the coefficients on Federal Reserve's Greenbook forecasts, $\beta_g$ in equation (9), estimated in the rolling regressions. The figure suggests that the coefficient averages around unity. However, the coefficient seems to have been decreasing over time over all horizons. For example, the bottom two panels depict the explanatory power of the Greenbook forecasts over that of the SPF's forecasts, and show that it clearly decreased over time. The picture also shows a mild revival of the information advantage around 1995-2001. The top two panels in the figure depict the explanatory power of the Greenbook over that of the BCEI forecasts; they reinforce the evidence in favor of the presence of additional explanatory power of the Greenbook forecasts around 1995, which starts diminishing around 2001.

INSERT FIGURE 6 HERE

# 5 High Frequency Forecasts

In this section we consider high frequency forecasts provided by the International Money Market Services (MMS) real-time database. The data includes median values from a weekly survey of 40 money managers on economic indicators whose values are expected to be announced/released during the following week. Balduzzi, Elton and Green (2001) and Urich and Wachtel (1984) have shown that based on traditional tests, forecast rationality cannot be rejected for most economic indicators. Our objective is to evaluate whether their findings are robust to instabilities. To do so, we implement our "Fluctuation Optimality" test, where as before we set $m = 60$.

The announcements that we consider include those regarding a variety of real and nominal economic fundamentals for the U.S. considered in Andersen et al. (2003) and described in more detail in Table 8. Since the economic indicators are released at different frequencies, the number of observations in each indicator varies. For example, initial jobless claims are released weekly, thus the MMS real-time database has weekly forecasts and realized observations for this indicator. On the contrary, changes in the non-farm payroll employment are released monthly implying fewer forecasted and realized observations for this variable. Table 8 also reports description of the data, including the source of the realized real-time values, the starting date of the series (the sample ends in January/February 2009 for most series) and the number of observations in our sample.

INSERT TABLE 8 HERE

Table 9 shows the results from both the traditional tests (column labeled "Traditional") as well our proposed "Fluctuation Rationality" test (column labeled "Fluctuation"). Asterisks denote significance at the 5% level. Both tests provide similar results for a variety of indicators. In what follows, we mainly focus on the indicators for which the implications of the two tests disagree. For housing starts, capacity utilization, consumer confidence, new home sales, consumer credit, GDP (annualized) preliminary release, CPI excluding Food and Energy Prices, as well as PPI excluding Food and Energy Prices, the "Traditional" test fails to reject forecast rationality, while the "Fluctuation" test does. In contrast, the rationality of the Advance Retail Sales, Business Inventories, GDP (annualized) advance release, and GDP Price Index/GDP Price Deflator (advanced) are rejected based on the traditional test, while rationality is not rejected based on a Fluctuation-type test. This result highlights the trade-offs between using traditional tests relative to our proposed "Fluctuation Optimality" tests: the former are more powerful when the data are stationary (as the Fluctuation-type test is based on rolling windows whose length is shorter than the total out-of-sample period), whereas the latter are more powerful in the presence of instabilities. Overall, our proposed test provides more empirical evidence against forecast rationality than previously reported.

INSERT TABLE 9 HERE

Closer investigation of the test statistics, depicted in Figure 7, shows that, for housing starts, capacity utilization, new home sales, PPI excluding Food and Energy Price (mom), the rejections in the Fluctuation Rationality test is driven by sporadic episodes in the beginning of 2000s. The case of consumer confidence is very interesting in that there has been a clearly observed reversion towards rationality since 2001. The forecast of annualized preliminary releases of GDP shows the opposite tendency: there is less evidence of forecast rationality since 2006. Consumer credit as well as the CPI excluding Food and Energy tend to have somewhat more prolonged periods away from rationality: the first one in the 2000s and the second from 1998 to 2004.

INSERT FIGURE 7 HERE

# 6 Conclusion

This paper proposes new forecast optimality tests that can be used in unstable environments. The tests we propose can be applied to test forecast unbiasedness, efficiency (e.g. Mincer and Zarnowitz's (1969) rationality regressions), encompassing, serial uncorrelation and, in general, any regression-based tests of forecasting ability. Our test statistics depends on a nuisance parameter and we tabulate the critical values of our test statistics as a function of that, making the test easily implementable. Our paper then analyzes the size properties of the test that we propose in small samples, as well as the power of our tests relative to traditional tests in the presence of instabilities. We show that traditional tests may fail to reject forecast optimality in the presence of instabilities whereas our test performs well in that regard.

In the first empirical analysis we compare various private sector forecasts to those of the Federal Reserve Greenbook. We reject the forecast rationality of all these forecasts at some point in time. However, even after allowing for time-variation, we find significant evidence in favor of the Fed's additional information advantage over the private sector when predicting future inflation. The second empirical analysis focuses on higher frequency forecasts from the Money Market Services in regards to a variety of economic fundamentals. We find that forecast rationality is frequently rejected by both traditional and Fluctuation-type tests, although our test overall uncovers more empirical evidence against rationality compared to the traditional tests.

# 7 References


Andersen, T., T. Bollerslev, F.X. Diebold and C. Vega (2003), "Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange", *American Economic Review* 93(1),38-62.

Balduzzi, P., E.J. Elton and T.C. Green (2001), "Economic News and Bond Prices: Evidence from the U.S. Treasury Market", *Journal of Financial and Quantitative Analysis* 36(4), 523-543.

Croushore, D. (1998), "Evaluating Inflation Forecasts", *Federal Reserve Bank of Philadelphia Working Paper* No 98-14.

Croushore, D. (2011), "Two Dimension of Forecast Evaluation: Vintages and Sub-Samples", *mimeo.*

Croushore, D. and T. Stark (2001), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics* 105(1), 111-130.

Diebold, F.X. and J.A. Lopez (1996), "Forecast Evaluation and Combination,"in G.S. Maddala and C.R. Rao (eds.), *Handbook of Statistics.* Amsterdam: North-Holland, 241-268.

Faust, J. and J. Wright (2009), "Comparing Greenbook and Reduced Form Forecasts using a Large Real-Time Dataset", *Journal of Business and Economic Statistics* 27, 468-479.

Giacomini, R. and B. Rossi (2010), "Forecast Comparisons in Unstable Environments", *Journal of Applied Econometrics* 25(4),595-620.

Granger, C.W.J. and P. Newbold (1986), *Forecasting Economic Time Series* (2nd ed.), New York: Academic Press.

Mincer, J. and V. Zarnowitz (1969), "The Evaluation of Economic Forecasts," in *Economic Forecasts and Expectations*, ed. J. Mincer, New York: National Bureau of Economic Research, 81–111.

Newey, W., and K. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica* 55, 703-708.

Patton, A. and A. Timmermann (2010), "Generalized Forecast Errors, A Change of Measure, and Forecast Optimality Conditions ", in *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, ed. T. Bollerslev, J.R. Russell, and M.W. Watson, Oxford University Press.

Patton, A. and A. Timmermann (2011), "Forecast Rationality Tests Based on Multi-Horizon Bounds", *Journal of Business and Economic Statistics*, forthcoming.

Romer, C.D. and D.H. Romer (2000), "Federal Reserve Information and the Behavior of Interest Rates", *The American Economic Review* 90(3), 429-457.

Rossi, B. (2011a), "Comment to: Forecast Rationality Tests Based on Multi-Horizon Bounds, by A. Patton and A. Timmermann", *Journal of Business and Economic Statistics*, forthcoming.

Rossi, B. (2011b), "Advances in Forecasting Under Instabilities", in: G. Elliott and A. Timmermann, *Handbook of Forecasting*, Vol. 2, Elsevier.

Urich, T.J and P. Wachtel (1984), "The Effects of Inflation and Money Supply Announcements on Interest Rates", *Journal of Finance* 39(4), 1177-88.

West, K.D. and M.W. McCracken (1998), "Regression Based Tests of Predictive Ability", *International Economic Review* 39(4), 816-840.

Wooldridge, J.M. and H. White (1988), "Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes", *Econometric Theory* 4(2), 210-230.

# Tables

**Table 1. Critical Values for the Fluctuation Optimality Test**

**Panel A. 10% Significance Level**

| | $\mu$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| q | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
| 1 | 10.5066 | 9.0503 | 8.0245 | 7.1035 | 6.3957 | 5.6112 | 5.1113 | 4.6141 | 3.9748 |
| 2 | 21.2392 | 18.0544 | 15.8290 | 13.9122 | 13.0720 | 11.1526 | 10.4549 | 9.0570 | 7.8723 |
| 3 | 31.4497 | 26.8866 | 23.7832 | 21.4577 | 19.6097 | 17.4180 | 15.3225 | 13.5010 | 11.4381 |
| 4 | 43.5150 | 36.9028 | 32.8187 | 28.4075 | 25.1774 | 23.3645 | 20.5785 | 17.6700 | 15.5384 |
| 5 | 52.4148 | 45.7998 | 39.6896 | 35.7848 | 32.0200 | 28.4850 | 26.2204 | 23.1738 | 19.1090 |
| 6 | 62.6771 | 54.3749 | 47.4711 | 42.4503 | 38.4920 | 34.9394 | 30.4063 | 27.9807 | 23.8787 |
| 7 | 74.8406 | 62.3659 | 56.2449 | 49.0721 | 44.4213 | 39.6189 | 36.4280 | 33.0852 | 26.9654 |
| 8 | 84.5728 | 72.8813 | 63.2267 | 56.8973 | 51.5069 | 45.7856 | 41.3975 | 36.7853 | 31.2008 |
| 9 | 94.9541 | 81.4986 | 72.0148 | 64.7598 | 57.1844 | 51.4226 | 46.4180 | 41.2775 | 33.9880 |
| 10 | 104.5994 | 90.8578 | 80.1290 | 69.7875 | 63.1391 | 58.7175 | 51.6605 | 45.565 | 38.6993 |
| 11 | 116.4526 | 98.2358 | 87.1706 | 79.0583 | 70.8704 | 63.5777 | 59.1131 | 51.1032 | 43.5269 |
| 12 | 126.3435 | 105.4528 | 95.2778 | 84.5975 | 76.9749 | 67.6842 | 63.2760 | 54.5424 | 48.1018 |
| 13 | 137.9833 | 118.9705 | 106.0615 | 91.2143 | 83.4945 | 75.5210 | 65.9104 | 60.1986 | 52.3220 |
| 14 | 149.0388 | 124.9307 | 111.3770 | 99.0350 | 90.1049 | 81.6688 | 72.7149 | 65.0909 | 53.7370 |
| 15 | 158.0203 | 134.0778 | 118.2550 | 104.8551 | 98.7227 | 87.4664 | 76.2884 | 69.3919 | 58.5763 |

**Panel B. 5% Significance Level**

| | $\mu$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| q | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
| 1 | 12.0846 | 10.5969 | 9.6550 | 8.7536 | 7.7583 | 6.961 | 6.4958 | 6.1299 | 5.3700 |
| 2 | 23.9304 | 21.0152 | 18.8106 | 16.9024 | 16.4506 | 14.5181 | 13.2906 | 11.9523 | 10.6583 |
| 3 | 35.8110 | 31.4406 | 28.0387 | 25.7908 | 24.5830 | 21.9864 | 19.7824 | 17.7214 | 15.1007 |
| 4 | 49.4366 | 43.1530 | 39.3683 | 34.2290 | 31.3434 | 29.3066 | 26.0227 | 22.9933 | 20.9018 |
| 5 | 59.4929 | 54.2582 | 46.9204 | 43.5296 | 40.2826 | 36.1054 | 33.2983 | 30.5946 | 25.6162 |
| 6 | 71.5833 | 63.6046 | 56.9141 | 51.3008 | 47.6269 | 43.5316 | 38.8591 | 36.5334 | 31.9232 |
| 7 | 85.5798 | 73.1198 | 67.0436 | 60.3735 | 54.9621 | 51.0055 | 46.4178 | 42.5572 | 37.3060 |
| 8 | 96.5994 | 82.8340 | 75.8613 | 69.3683 | 64.7512 | 57.4847 | 54.7416 | 48.6285 | 40.4862 |
| 9 | 107.9342 | 95.6691 | 85.9190 | 79.9041 | 70.2681 | 64.2407 | 60.5165 | 53.4894 | 46.1068 |
| 10 | 120.5426 | 107.9044 | 95.7044 | 84.9419 | 78.7060 | 71.8195 | 65.8906 | 59.5306 | 51.3579 |
| 11 | 132.6715 | 114.3937 | 105.5542 | 95.1193 | 88.5071 | 78.8749 | 74.9576 | 64.6169 | 59.0124 |
| 12 | 145.6436 | 124.7966 | 114.8308 | 101.9825 | 93.2882 | 87.7456 | 79.7758 | 70.5897 | 63.5020 |
| 13 | 155.9087 | 139.2837 | 126.6474 | 110.2505 | 101.8596 | 97.2653 | 82.8943 | 78.5014 | 69.1920 |
| 14 | 169.6245 | 144.8327 | 130.6773 | 119.2178 | 111.8770 | 101.7223 | 92.7027 | 84.8191 | 71.8956 |
| 15 | 180.4920 | 157.8693 | 141.1974 | 130.0529 | 120.9388 | 110.0675 | 95.9919 | 88.8136 | 78.0180 |

**Panel C. 1% Significance Level**

| | $\mu$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| q | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
| 1 | 15.2034 | 14.1709 | 12.8862 | 12.5024 | 11.3655 | 10.2615 | 9.4840 | 9.4669 | 8.8229 |
| 2 | 30.8516 | 28.2962 | 25.9860 | 24.0338 | 23.3590 | 21.7058 | 20.1484 | 18.3749 | 16.6201 |
| 3 | 45.5649 | 41.8301 | 38.5925 | 35.8044 | 34.0350 | 32.5569 | 30.2612 | 27.1766 | 23.8648 |
| 4 | 62.9754 | 54.9380 | 55.6239 | 48.0043 | 45.2153 | 44.5599 | 41.1181 | 35.1465 | 31.2726 |
| 5 | 78.8954 | 69.0344 | 63.8089 | 59.8863 | 56.6435 | 53.3724 | 50.3065 | 48.5804 | 41.4091 |
| 6 | 91.9499 | 84.6385 | 76.9111 | 73.6821 | 67.9723 | 63.2696 | 59.3231 | 55.8028 | 49.0648 |
| 7 | 112.8762 | 96.4070 | 91.6098 | 86.1555 | 77.8255 | 75.7540 | 69.5349 | 63.2412 | 59.2658 |
| 8 | 125.4023 | 109.8004 | 105.4471 | 96.4564 | 98.2139 | 84.6103 | 84.8680 | 77.9564 | 67.3690 |
| 9 | 140.4446 | 126.4322 | 119.2996 | 109.7265 | 101.7085 | 92.3127 | 91.4050 | 79.0176 | 70.5083 |
| 10 | 153.1288 | 145.7988 | 132.9398 | 121.6310 | 113.1604 | 107.4479 | 98.5991 | 92.6189 | 84.2430 |
| 11 | 170.3148 | 151.9902 | 143.2056 | 130.4234 | 127.8291 | 117.4778 | 114.4482 | 100.9009 | 94.5669 |
| 12 | 186.9767 | 167.1092 | 153.7950 | 146.5000 | 127.5479 | 130.2217 | 117.6577 | 112.7820 | 97.8878 |
| 13 | 205.5454 | 181.7725 | 171.6468 | 151.0495 | 149.8351 | 143.3378 | 131.0534 | 120.8815 | 109.4559 |
| 14 | 217.3423 | 195.1713 | 177.0460 | 162.8551 | 155.2296 | 151.6191 | 146.8675 | 128.3316 | 116.6285 |
| 15 | 231.7324 | 210.6466 | 186.2040 | 187.4957 | 168.7556 | 159.4214 | 144.4295 | 144.6111 | 124.8982 |

Note. The table reports $\kappa_{\alpha,q}$, the critical values at 10%, 5%, and 1% significance levels respectively, for $max_t\ \mathcal{W}_{t,m}$ for various values of $\mu = [m/P]$ and number of restrictions, $q$.

**Table 2: Size. Recursive Estimation**

| | Traditional Test | | Fluctuation Test | |
|---|---|---|---|---|
| R/P | 100 | 200 | 100 | 200 |
| **Panel A. Mean Prediction Error** | | | | |
| 25 | 0.0486 | 0.0496 | 0.0500 | 0.0648 |
| 50 | 0.0454 | 0.0494 | 0.0522 | 0.0630 |
| 100 | 0.0572 | 0.0482 | 0.0582 | 0.0678 |
| 200 | 0.0468 | 0.0520 | 0.0580 | 0.0712 |
| 300 | 0.0534 | 0.0454 | 0.0526 | 0.0624 |
| 400 | 0.0508 | 0.0538 | 0.0546 | 0.0632 |
| **Panel B. Efficiency Test** | | | | |
| 25 | 0.0498 | 0.0544 | 0.0686 | 0.0832 |
| 50 | 0.0542 | 0.0478 | 0.0644 | 0.0874 |
| 100 | 0.0546 | 0.0518 | 0.0726 | 0.0842 |
| 200 | 0.0512 | 0.0502 | 0.0670 | 0.0812 |
| 300 | 0.0570 | 0.0546 | 0.0658 | 0.0820 |
| 400 | 0.0566 | 0.0484 | 0.0694 | 0.0800 |
| **Panel C. Encompassing Tests** | | | | |
| 25 | 0.0932 | 0.0830 | 0.0980 | 0.1050 |
| 50 | 0.0784 | 0.0702 | 0.0828 | 0.0928 |
| 100 | 0.0656 | 0.0602 | 0.0772 | 0.0786 |
| 200 | 0.0530 | 0.0584 | 0.0570 | 0.0736 |
| 300 | 0.0560 | 0.0538 | 0.0680 | 0.0766 |
| 400 | 0.0546 | 0.0566 | 0.0594 | 0.0628 |
| **Panel D. Serial (Un)Correlation Tests** | | | | |
| 25 | 0.0464 | 0.0488 | 0.0424 | 0.0570 |
| 75 | 0.0470 | 0.0486 | 0.0458 | 0.0542 |
| 100 | 0.0552 | 0.0486 | 0.0476 | 0.0528 |
| 200 | 0.0554 | 0.0448 | 0.0450 | 0.0486 |
| 300 | 0.0492 | 0.0548 | 0.0484 | 0.0516 |
| 400 | 0.0498 | 0.0500 | 0.0486 | 0.0586 |

Note. Tables 2 and 3 report empirical rejection frequencies of the test statistics $max_t \ \mathcal{W}_{t,m}$ (column labeled "Fluctuation Test") as well as the traditional test statistics $\mathrm{t}_{\alpha}$ and $\mathrm{t}_{\alpha}^{c}$ (column labeled "Traditional Test") under the recursive and rolling estimation schemes respectively (see DGP in Section 3 for details). m = 50. Nominal size is 0.05.

**Table 3. Size. Rolling Window Estimation**

**Panel A. Mean Prediction Error**

| | I. Correction | | II. No Correction | |
|---|---|---|---|---|
| | Traditional Test | | | |
| R/P | 100 | 200 | 100 | 200 |
| 25 | 0.4014 | 0.5540 | 0.0368 | 0.0402 |
| 50 | 0.2554 | 0.4216 | 0.0410 | 0.0444 |
| 100 | 0.1162 | 0.2598 | 0.0542 | 0.0482 |
| 200 | 0.0536 | 0.1132 | 0.0446 | 0.0484 |
| 300 | 0.0548 | 0.0658 | 0.0516 | 0.0452 |
| 400 | 0.0538 | 0.0632 | 0.0510 | 0.0540 |
| | Fluctuation Test | | | |
| 25 | 0.4686 | 0.6668 | 0.0278 | 0.0224 |
| 50 | 0.1436 | 0.1992 | 0.0366 | 0.0406 |
| 100 | 0.0686 | 0.0830 | 0.0518 | 0.0608 |
| 200 | 0.0608 | 0.0764 | 0.0584 | 0.0716 |
| 300 | 0.0564 | 0.0630 | 0.0550 | 0.0610 |
| 400 | 0.0536 | 0.0622 | 0.0524 | 0.0614 |

**Panel B. Efficiency Tests**

| | I. Correction | | II. No Correction | |
|---|---|---|---|---|
| | Traditional Test | | | |
| R/P | 100 | 200 | 100 | 200 |
| 25 | 0.5678 | 0.9342 | 0.0224 | 0.0270 |
| 50 | 0.1468 | 0.3388 | 0.0150 | 0.0042 |
| 100 | 0.0748 | 0.0998 | 0.0270 | 0.0050 |
| 200 | 0.0546 | 0.0582 | 0.0440 | 0.0212 |
| 300 | 0.0604 | 0.0560 | 0.0546 | 0.0392 |
| 400 | 0.0552 | 0.0472 | 0.0532 | 0.0394 |
| | Fluctuation Test | | | |
| 25 | 0.5332 | 0.7738 | 0.0526 | 0.0598 |
| 50 | 0.1504 | 0.2320 | 0.0458 | 0.0602 |
| 100 | 0.0854 | 0.1046 | 0.0654 | 0.0786 |
| 200 | 0.0666 | 0.0866 | 0.0638 | 0.0828 |
| 300 | 0.0696 | 0.0854 | 0.0658 | 0.0820 |
| 400 | 0.0710 | 0.0834 | 0.0698 | 0.0816 |

| Panel C. Encompassing Tests | | | | |
|---|---|---|---|---|
| | I. Correction | | II. No Correction | |
| | Traditional Test | | | |
| R/P | 100 | 200 | 100 | 200 |
| 25 | 0.1576 | 0.2286 | 0.1348 | 0.2022 |
| 50 | 0.1036 | 0.1236 | 0.0898 | 0.1046 |
| 100 | 0.0850 | 0.0848 | 0.0748 | 0.0702 |
| 200 | 0.0614 | 0.0688 | 0.0538 | 0.0606 |
| 300 | 0.0604 | 0.0622 | 0.0538 | 0.0534 |
| 400 | 0.0626 | 0.0578 | 0.0558 | 0.0552 |
| | Fluctuation Test | | | |
| 25 | 0.1388 | 0.1786 | 0.1258 | 0.1688 |
| 50 | 0.1142 | 0.1508 | 0.1008 | 0.1316 |
| 100 | 0.0984 | 0.1020 | 0.0858 | 0.0956 |
| 200 | 0.0626 | 0.0826 | 0.0592 | 0.0774 |
| 300 | 0.0682 | 0.0842 | 0.0670 | 0.0780 |
| 400 | 0.0614 | 0.0688 | 0.0594 | 0.0660 |

| Panel D. Serial (Un)Correlation Tests | | | | |
|---|---|---|---|---|
| | I. Correction | | II. No Correction | |
| | Traditional Test | | | |
| R/P | 100 | 200 | 100 | 200 |
| 25 | 0.0902 | 0.1442 | 0.0036 | 0.0038 |
| 50 | 0.0576 | 0.0716 | 0.0056 | 0.0028 |
| 100 | 0.0530 | 0.0508 | 0.0272 | 0.0054 |
| 200 | 0.0474 | 0.0468 | 0.0460 | 0.0220 |
| 300 | 0.0498 | 0.0506 | 0.0470 | 0.0424 |
| 400 | 0.0552 | 0.0534 | 0.0486 | 0.0430 |
| | Fluctuation Test | | | |
| 25 | 0.0918 | 0.1064 | 0.0042 | 0.0024 |
| 50 | 0.0656 | 0.0860 | 0.0200 | 0.0210 |
| 100 | 0.0692 | 0.0674 | 0.0432 | 0.0458 |
| 200 | 0.0534 | 0.0692 | 0.0428 | 0.0464 |
| 300 | 0.0594 | 0.0750 | 0.0464 | 0.0510 |
| 400 | 0.0552 | 0.0640 | 0.0472 | 0.0552 |

**Table 4: Power Analysis**

| Panel A. DGP 1 - Mean Forecast Error | | |
|---|---|---|
| $b$ | I. Traditional | II. Fluctuation |
| 0 | 0.0482 | 0.0448 |
| 0.1 | 0.0486 | 0.0504 |
| 0.2 | 0.0460 | 0.0676 |
| 0.3 | 0.0432 | 0.0848 |
| 0.4 | 0.0350 | 0.1014 |
| 0.5 | 0.0312 | 0.1358 |
| 0.6 | 0.0222 | 0.1456 |
| 0.7 | 0.0214 | 0.1740 |
| 0.8 | 0.0138 | 0.1870 |
| 0.9 | 0.0086 | 0.1826 |
| 1.0 | 0.0044 | 0.1886 |
| **Panel B. DGP 2 - Efficiency** | | |
| $b$ | I. Traditional | II. Fluctuation |
| 0 | 0.0540 | 0.0490 |
| 0.50 | 0.0680 | 0.0998 |
| 1.00 | 0.0362 | 0.2152 |
| 1.50 | 0.0110 | 0.4110 |
| 2.00 | 0.0016 | 0.6438 |
| 2.50 | 0 | 0.8278 |
| 3.00 | 0 | 0.9260 |
| 3.50 | 0 | 0.9782 |
| 4.00 | 0 | 0.9922 |
| 4.50 | 0 | 0.9980 |
| 5.00 | 0 | 1.0000 |

| Panel C. DGP 2 - Encompassing | | |
|---|---|---|
| $b$ | I. Traditional | II. Fluctuation |
| 0 | 0.0576 | 0.0542 |
| 0.50 | 0.0542 | 0.0476 |
| 1.00 | 0.0590 | 0.0394 |
| 1.50 | 0.0292 | 0.0466 |
| 2.00 | 0.0066 | 0.0686 |
| 2.50 | 0.0002 | 0.0996 |
| 3.00 | 0 | 0.1624 |
| 3.50 | 0 | 0.2362 |
| 4.00 | 0 | 0.3364 |
| 4.50 | 0 | 0.4218 |
| 5.00 | 0 | 0.5086 |
| **Panel D. DGP 3 - Serial (Un)Correlation** | | |
| $b$ | I. Traditional | II. Fluctuation |
| 0 | 0.0512 | 0.0466 |
| 0.05 | 0.0522 | 0.0510 |
| 0.10 | 0.0558 | 0.0794 |
| 0.15 | 0.0664 | 0.1486 |
| 0.20 | 0.0802 | 0.2280 |
| 0.25 | 0.0930 | 0.3490 |
| 0.30 | 0.1216 | 0.4930 |
| 0.35 | 0.1484 | 0.6550 |
| 0.40 | 0.1760 | 0.7852 |
| 0.45 | 0.2120 | 0.8958 |
| 0.50 | 0.2418 | 0.9476 |

Note. The table reports empirical rejection frequencies of the test statistics $max_t\ \mathcal{W}_{t,m}$ (column labeled "Fluctuation Test") as well as the traditional test statistics $\mathrm{t}_\alpha$ and $\mathrm{t}_\alpha^c$ (column labeled "Traditional Test") under the recursive estimation scheme in the presence of time-variation. (see DGP 1, 2, and 3 in Section 3 for details). R = 300, P = 100, m = 50. Nominal size is 0.05.

**Table 5. MSFE Comparisons**

| Horizon | Greenbook | BCEI | SPF |
|---|---|---|---|
| Sample Start Date: | 1968:IV | 1980:I | 1968:IV |
| 0 | 1.06 | 0.97 | 1.32 |
| 1 | 1.76 | 1.29 | 2.28 |
| 2 | 2.33 | 1.56 | 3.02 |
| 3 | 2.47 | 1.92 | 3.67 |
| 4 | 2.71 | 2.31 | 4.30 |
| 5 | 2.70 | 2.61 | - - |

Note. MSFE is calculated as $(\pi_{t+h} - \hat{\pi}_{t+h,\ t})^2$ for various forecast horizons $h$.

**Table 6. Inflation Forecast Rationality Tests**

| Horizon | N. Obs. | Fluctuation | Traditional |
|---|---|---|---|
| | | Greenbook | |
| 0 | 149 | 39.64* | 0.15 |
| 1 | 149 | 46.89* | 0.65 |
| 2 | 143 | 49.41* | 0.33 |
| 3 | 134 | 41.59* | 0.02 |
| 4 | 109 | 37.89* | 0.01 |
| 5 | 74 | 91.96* | 3.33 |
| | | BCEI | |
| 0 | 100 | 43.74* | 11.95* |
| 1 | 100 | 51.97* | 15.69* |
| 2 | 100 | 74.22* | 22.38* |
| 3 | 98 | 135.79* | 44.67* |
| 4 | 74 | 167.51* | 64.54* |
| | | SPF | |
| 0 | 149 | 45.11* | 0.13 |
| 1 | 149 | 66.36* | 0.16 |
| 2 | 148 | 77.84* | 0.07 |
| 3 | 145 | 158.77* | 0.46 |

Note. The table reports the Traditional and Fluctuation Wald test statistics, $\mathcal{W}_P$ and $\max_t \mathcal{W}_{t,m}$, respectively. The traditional tests are based on the indicated number of observations. The "Fluctuation test" results are based on $m = 60$. The significance of the test statistics at the 5% significance level is indicated by $*$.

**Table 7. Fed's Information Advantage Over Private Sector's Forecasts**

| Horizon | N. Obs. | Fluctuation | Traditional |
|---|---|---|---|
| | | BCEI | |
| 0 | 100 | 49.81* | 14.11* |
| 1 | 100 | 93.02* | 34.87* |
| 2 | 100 | 60.00* | 20.93* |
| 3 | 98 | 28.10* | 11.05* |
| | | SPF | |
| 0 | 149 | 39.71* | 36.96* |
| 1 | 149 | 47.68* | 18.78* |
| 2 | 142 | 38.46* | 19.29* |
| 3 | 134 | 51.07* | 34.22* |

Note. The table reports the Traditional and Fluctuation Wald test statistics ($W_P$and $\max_t W_{t,m}$,respectively) for their respective null hypotheses. The traditional tests are based on the indicated number of observations. The Fluctuation test results are based on $m = 60$. The significance of the test statistics at the 5% significance level is indicated by $*$.

**Table 8. Data Description for MMS Database**

| | Source | Starting date | N. Obs. |
|---|---|---|---|
| Unemployment Rate | BLS | 1/5/90 | 230 |
| Consumer Price Index | BLS | 1/18/90 | 229 |
| Durable Goods Orders | BC | 1/26/90 | 228 |
| Housing Starts | BC | 1/18/90 | 228 |
| Leading Indicators | CB | 1/31/90 | 229 |
| Trade Balance | BEA | 1/17/90 | 227 |
| Change in Nonfarm Payrolls | BLS | 1/5/90 | 230 |
| Producer Price Index | BLS | 1/12/90 | 229 |
| Advance Retail Sales | BC | 1/12/90 | 229 |
| Capacity Utilization | FRB | 1/17/90 | 229 |
| Industrial Production | FRB | 1/17/90 | 229 |
| Business Inventories | BC | 1/16/90 | 229 |
| Construction Spending MoM | FMS | 2/1/90 | 228 |
| Consumer Confidence | CB | 7/30/91 | 211 |
| Factory Orders | BC | 1/5/90 | 230 |
| NAPM/ISM Manufacturing | NAPM | 2/1/90 | 229 |
| New Home Sales | BC | 1/3/90 | 228 |
| Personal Consumption | BEA | 1/29/90 | 228 |
| Personal Income | BEA | 1/29/90 | 226 |
| Monthly Budget Statement | TD | 2/22/90 | 227 |
| Consumer Credit | FRB | 1/8/90 | 229 |
| Initial Jobless Claims | LD | 7/18/91 | 911 |
| GDP Annualized Advanced | BEA | 1/26/90 | 77 |
| GDP Annualized Preliminary | BEA | 2/28/90 | 75 |
| GDP Annualized Final | BEA | 3/28/90 | 76 |
| CPI Ex Food and Energy MoM | BLS | 1/16/92 | 203 |
| PPI Ex Food and Energy MoM | BLS | 1/9/92 | 204 |
| Average Hourly Earnings MoM | BLS | 2/2/90 | 227 |
| Retail Sales Less Autos | BC | 1/12/90 | 228 |
| GDP P. Index/GDP Price Defl. A | BEA | 10/29/91 | 70 |
| GDP P. Index/GDP Price Defl. P | BEA | 12/4/91 | 68 |
| GDP P. Index/GDP Price Defl. F | BEA | 12/20/91 | 68 |

Note. BC - Census Bureau, BEA - Bureau of Economic Analysis, BLS - Bureau of Labor Statistics, CB - Conference Board, FRB - Federal Reserve Board of Governors, LD - Dept. of Labor, TD - Treasury Dept.

**Table 9. MMS Forecast Rationality Tests**

| Variable | Fluctuation | Traditional |
|---|---|---|
| Unemployment Rate | 51.02* | 8.89* |
| Consumer Price Index | 64.17* | 19.16* |
| Durable Goods Orders | 28.37* | 33.14* |
| Housing Starts | 30.38* | 5.88 |
| Leading Indicators | 45.11* | 34.80* |
| Trade Balance | 8.16 | 0.74 |
| Change in Nonfarm Payrolls | 18.39* | 21.18* |
| Producer Price Index | 94.82* | 11.87* |
| Advance Retail Sales | 12.13 | 8.65* |
| Capacity Utilization | 20.24* | 1.42 |
| Industrial Production | 27.35* | 15.65* |
| Business Inventories | 15.04 | 9.96* |
| Construction Spending MoM | 6.44 | 5.47 |
| Consumer Confidence | 18.89* | 2.31 |
| Factory Orders | 20.74* | 10.29* |
| NAPM/ISM Manufacturing | 4.27 | 0.81 |
| New Home Sales | 29.66* | 4.61 |
| Personal Consumption | 43.27* | 15.07* |
| Personal Income | 25.52* | 8.38* |
| Monthly Budget Statement | 39.10* | 19.11* |
| Consumer Credit | 23.07* | 2.03 |
| Initial Jobless Claims | 50.91* | 7.24* |
| GDP Annualized Advanced | 7.86 | 7.34* |
| GDP Annualized Preliminary | 20.01* | 3.32 |
| GDP Annualized Final | 3.29 | 0.69 |
| CPI Ex Food and Energy MoM | 27.45* | 2.20 |
| PPI Ex Food and Energy MoM | 16.81* | 3.69 |
| Average Hourly Earnings MoM | 6.69 | 1.21 |
| Retail Sales Less Autos | 20.39* | 20.72* |
| GDP P. Index/GDP Price Defl. A | 6.31 | 9.30* |
| GDP P. Index/GDP Price Defl. P | 3.07 | 3.19 |
| GDP P. Index/GDP Price Defl. F | 24.45* | 21.31* |

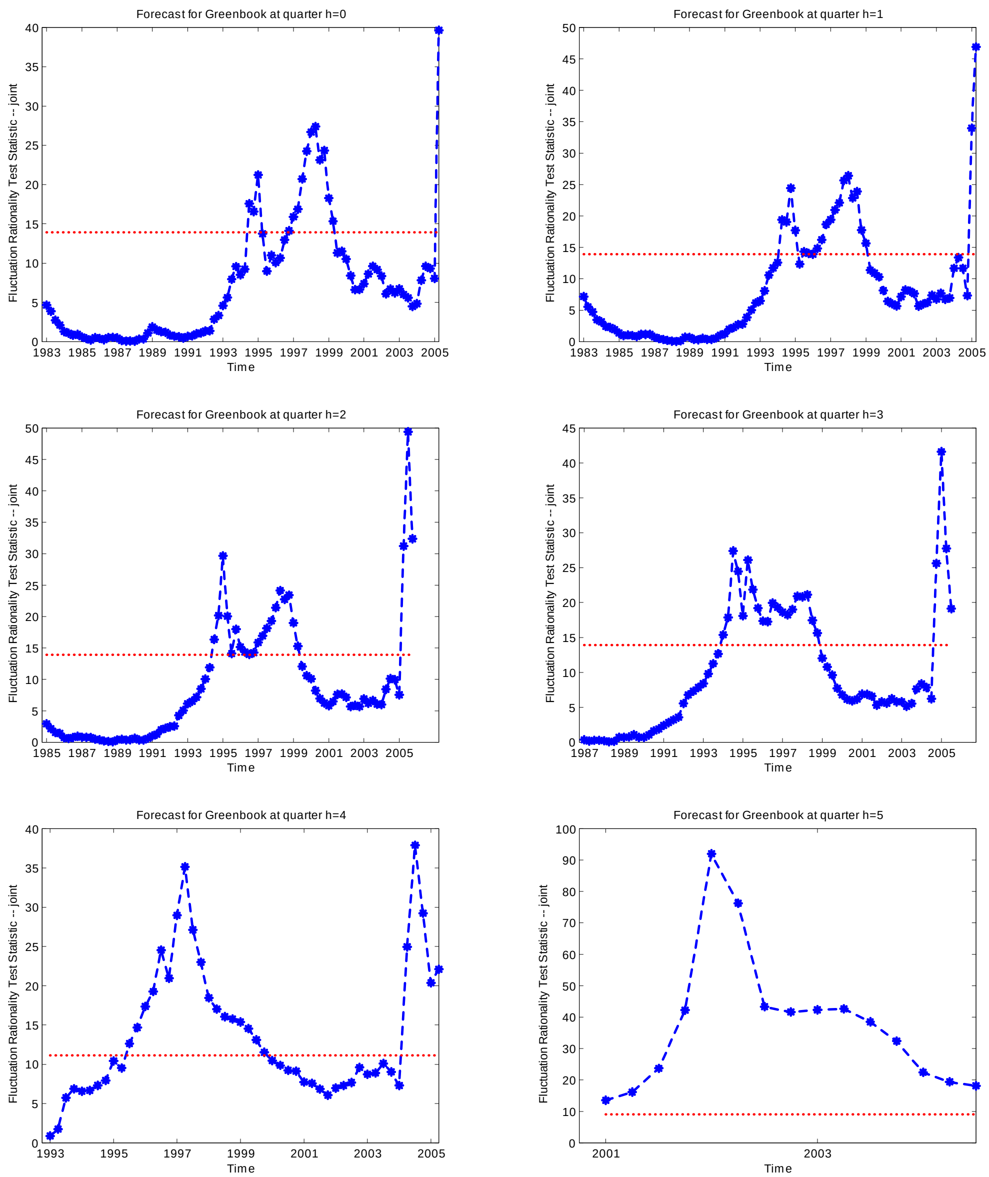
Note. The table reports the Traditional and Fluctuation Wald test statistics ($\mathcal{W}_P$ and $\max_t \mathcal{W}_{t,\ m}$, respectively). $m = 60$. The significance of the test statistics at the 5% significance level is indicated by $*$.

# Figures

## Figure 1: Inflation Forecasts

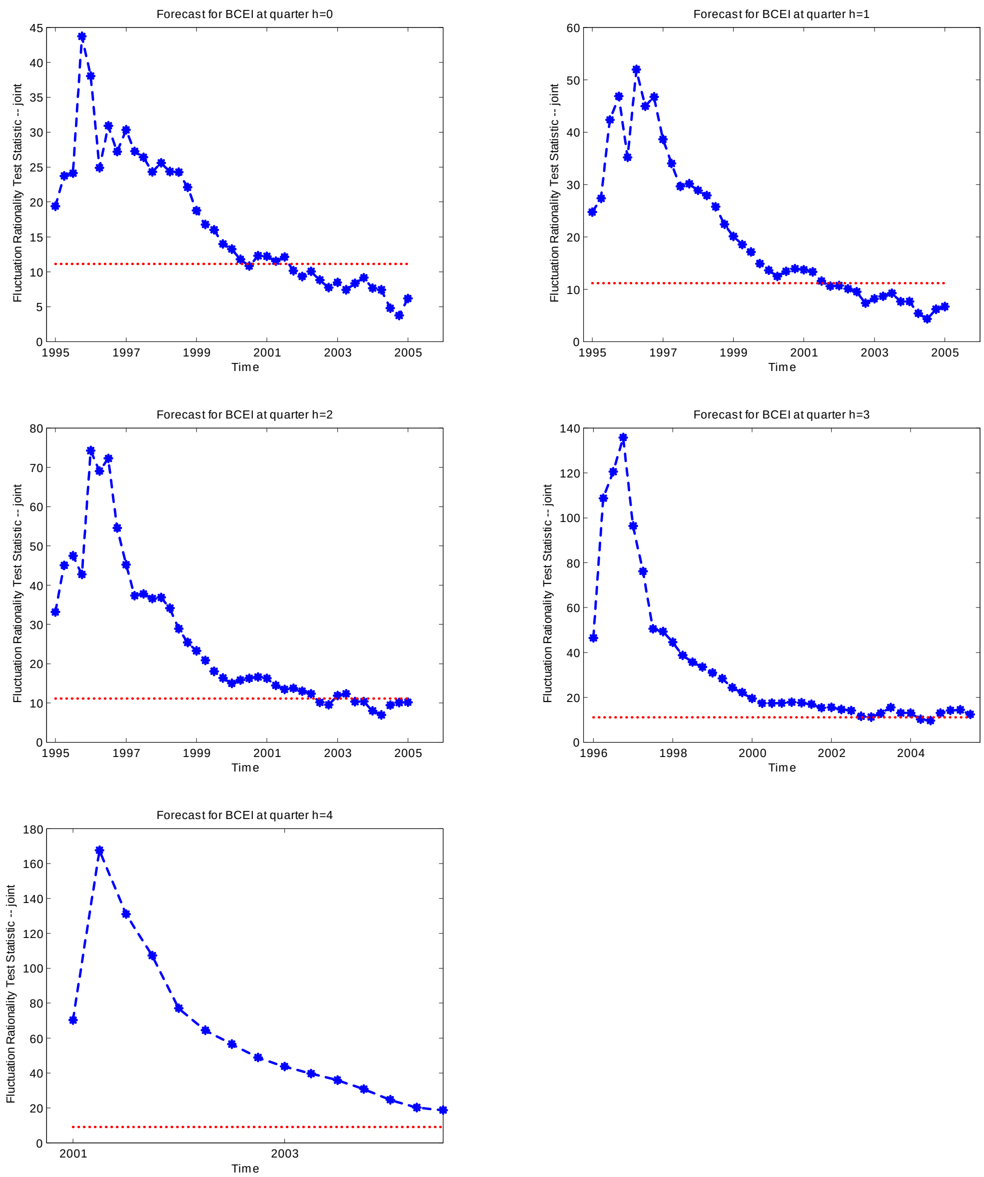

Note. The figure plots Greenbook, SPF, and BCEI forecasts of inflation for various forecast horizons $h$ in conjunction with the realized values of inflation for the corresponding horizon. If a forecast for a specific horizon by the corresponding agency does not exist, it is depicted as a missing value.

## Figure 2: Fluctuation Optimality Test for Greenbook Forecasts



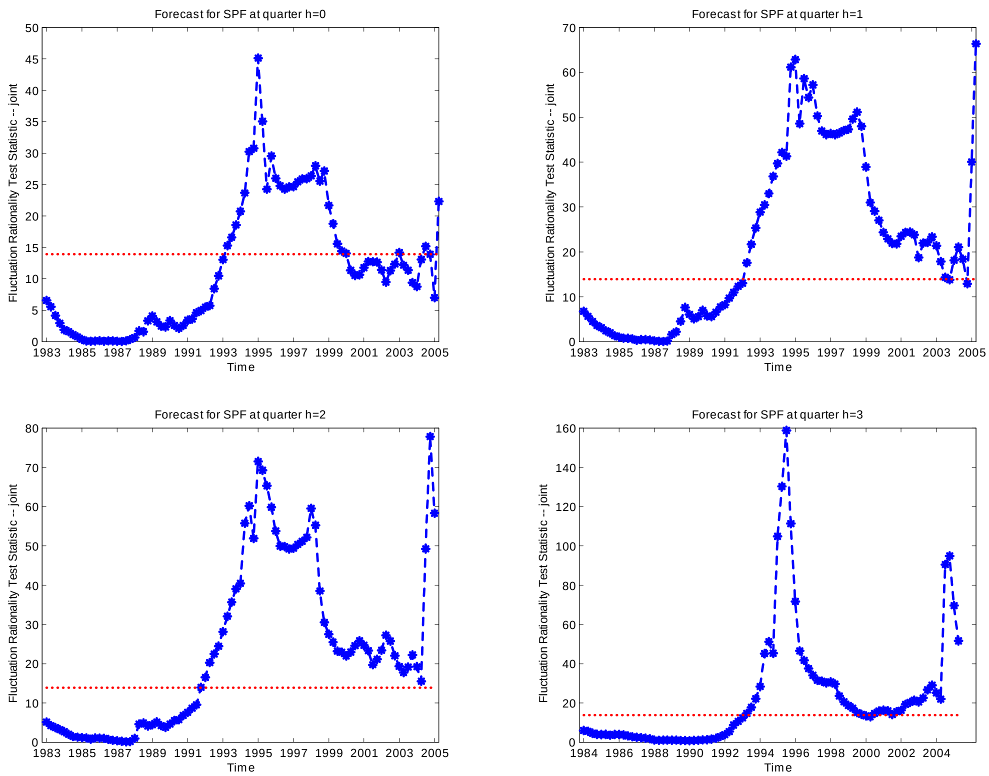

Note. The figure reports the time path of the test statistics $\mathcal{W}_{t,m}$ for the null hypothesis of forecast rationality under a recursive estimation scheme. $m = 60$ and the dotted line ("...") corresponds to the critical value at 5% significance level. If the test statistic is above the dotted line, we reject the null hypothesis of rationality at any point in time. The dates in the horizontal axis suggest a particular break-date.

# Figure 3: Fluctuation Optimality Test for BCEI Forecasts



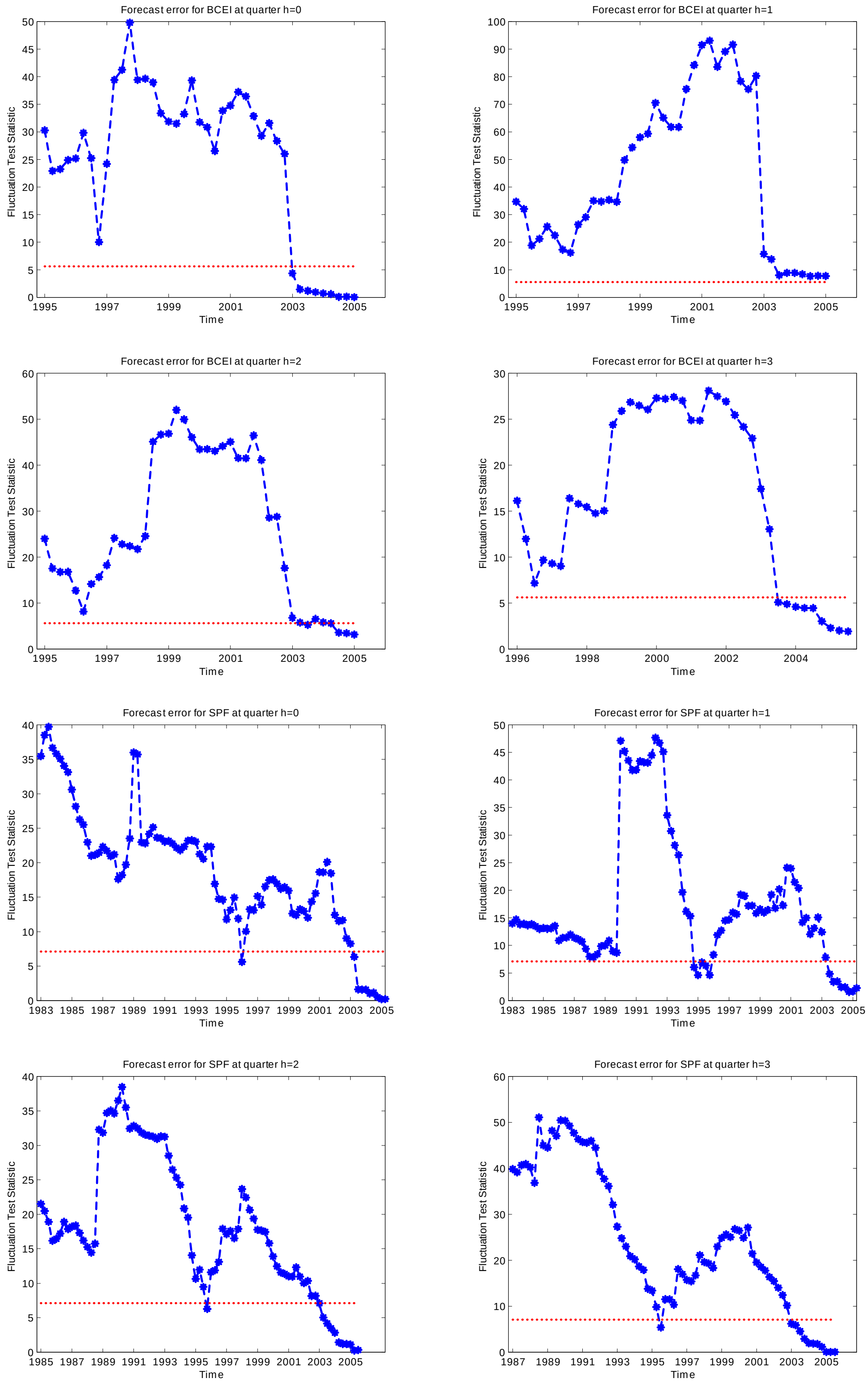

Note. The figure reports the time path of the test statistics $\mathcal{W}_{t,m}$ for the null hypothesis of forecast rationality under a recursive estimation scheme. $m = 60$ and the dotted line ("...") corresponds to the critical value at 5% significance level. If the test statistic is above the dotted line, we reject the null hypothesis of rationality at any point in time. The dates in the horizontal axis suggest a particular break-date.

# Figure 4: Fluctuation Optimality Test for SPF Forecasts



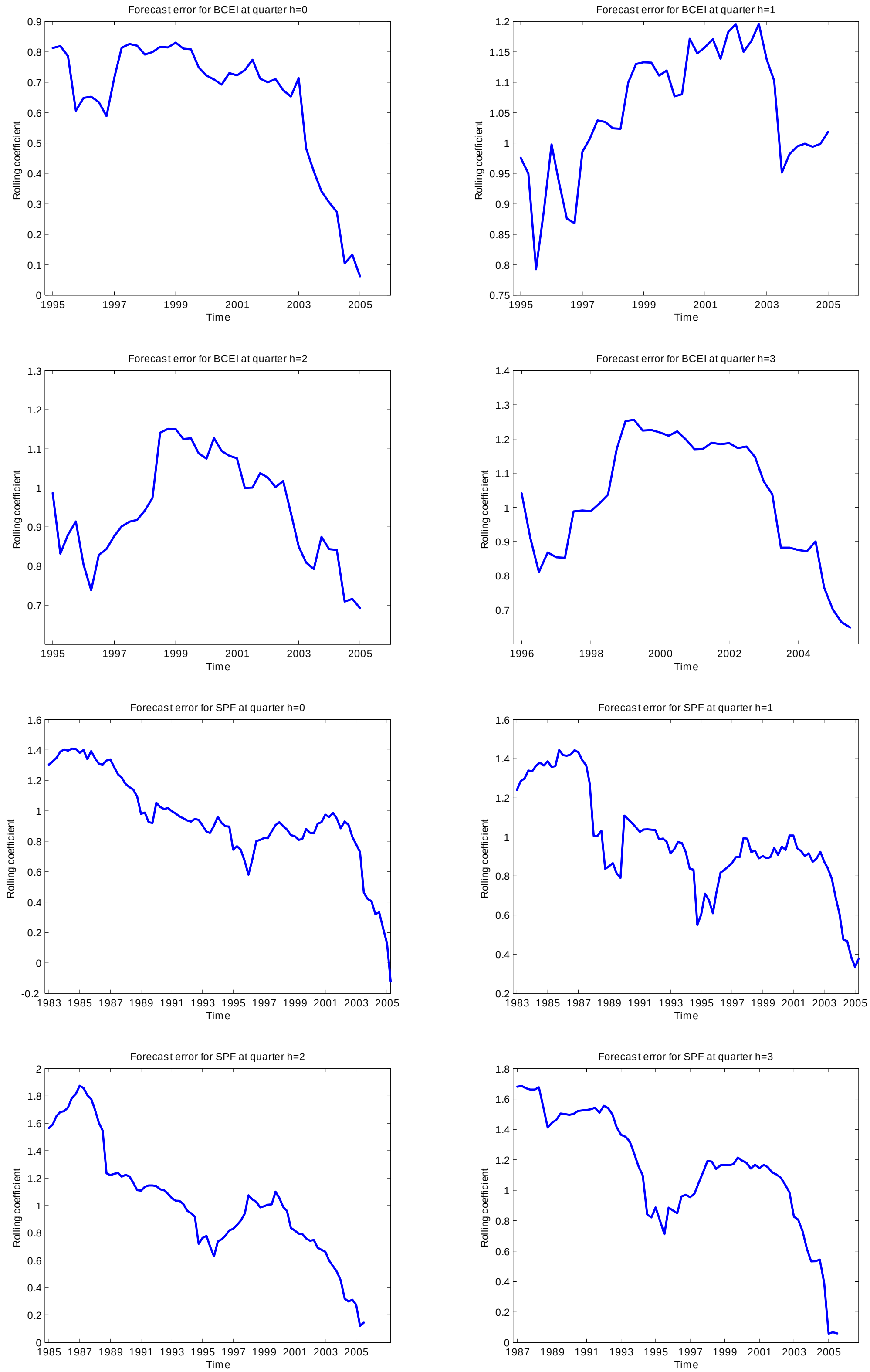

Note. The figure reports the time path of the test statistics $\mathcal{W}_{t,m}$ for the null hypothesis of forecast rationality under a recursive estimation scheme. $m = 60$ and the dotted line ("...") corresponds to the critical value at 5% significance level. If the test statistic is above the dotted line, we reject the null hypothesis of rationality at any point in time. The dates in the horizontal axis suggest a particular break-date.

# Figure 5: Fed's Informational Advantage



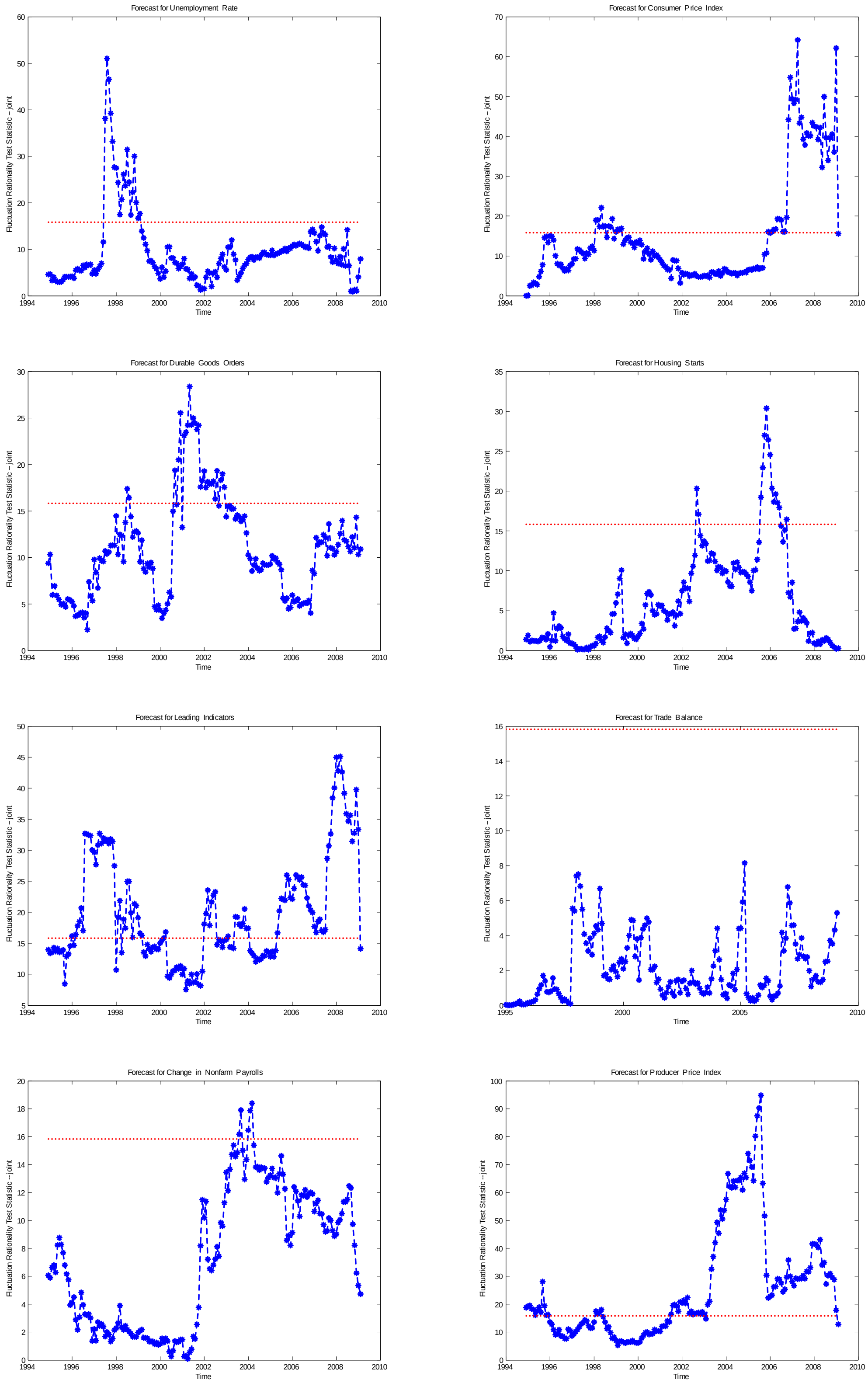

Note. The figure reports the test statistics $\mathcal{W}_{t,m}$ for the null hypothesis $E_t(\widehat{\beta}_{g,t}) = 0$ over time. $m = 60$.

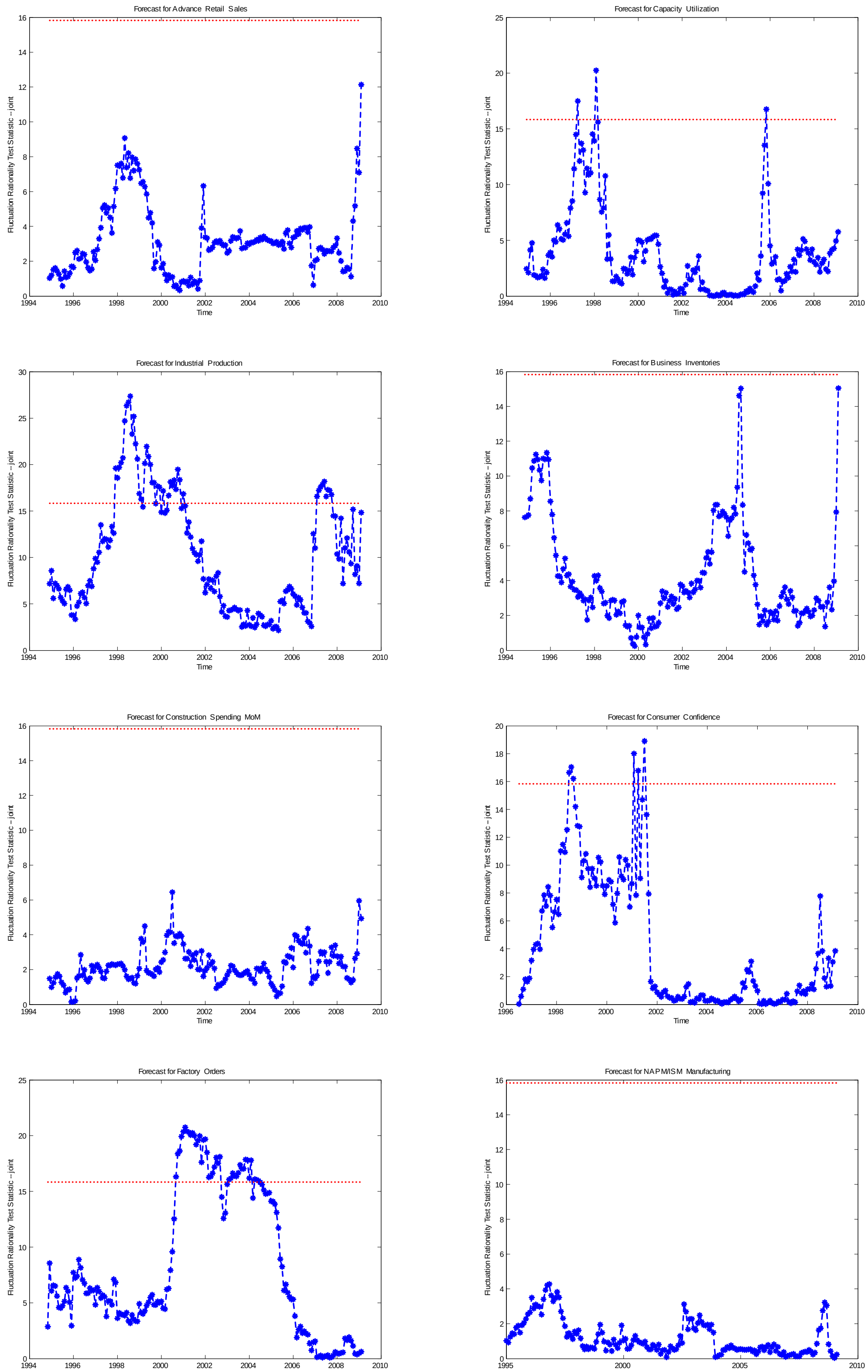
## Figure 6: Fed's Informational Coefficients



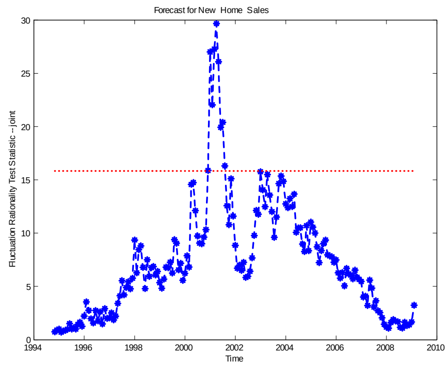

Note. The figure shows the rolling estimate of $\beta_g$ as in equation (9) based on $m = 60$.

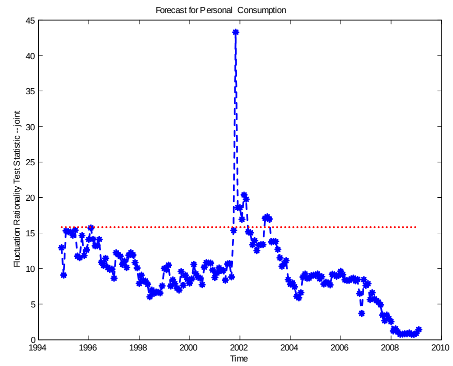
# Figure 7: Fluctuation Optimality Test for MMS Forecasts

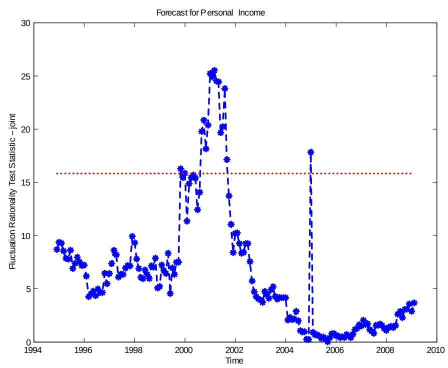


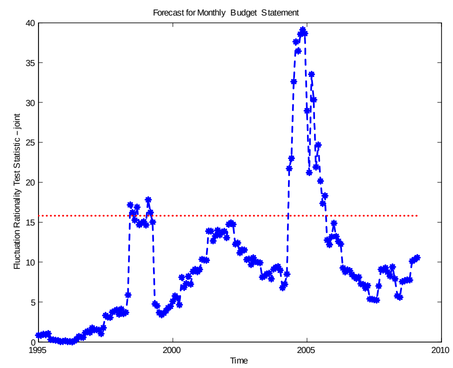


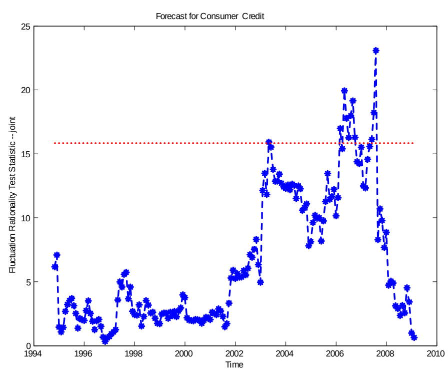
Forecast for Advance Retail Sales
Forecast for Capacity Utilization
Forecast for Industrial Production
Forecast for Business Inventories
Forecast for Construction Spending MoM
Forecast for Consumer Confidence
Forecast for Factory Orders
Forecast for NAPM/ISM Manufacturing
Fluctuation Rationality Test Statistic – joint
Time

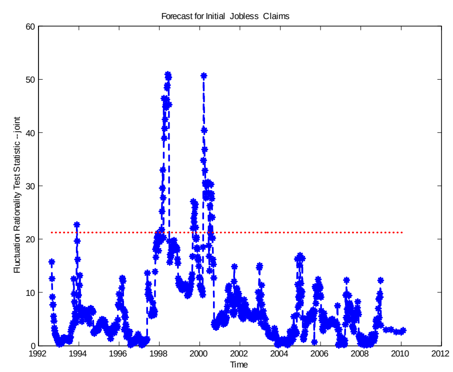

Forecast for New Home Sales
Fluctuation Rationality Test Statistic – joint
Time

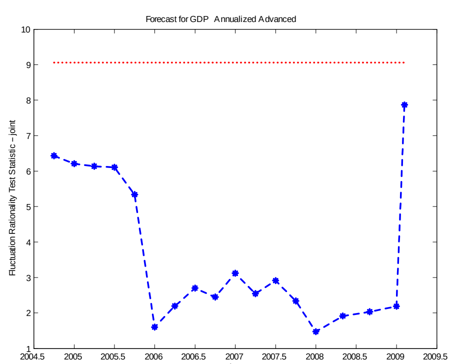

Forecast for Personal Consumption
Fluctuation Rationality Test Statistic – joint
Time

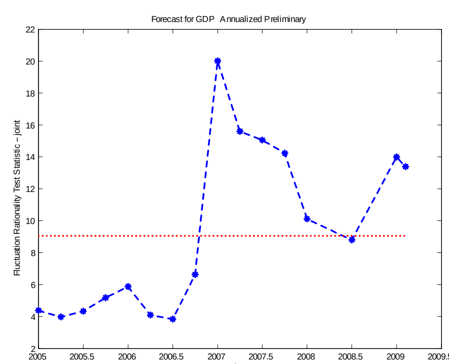

Forecast for Personal Income
Fluctuation Rationality Test Statistic – joint
Time

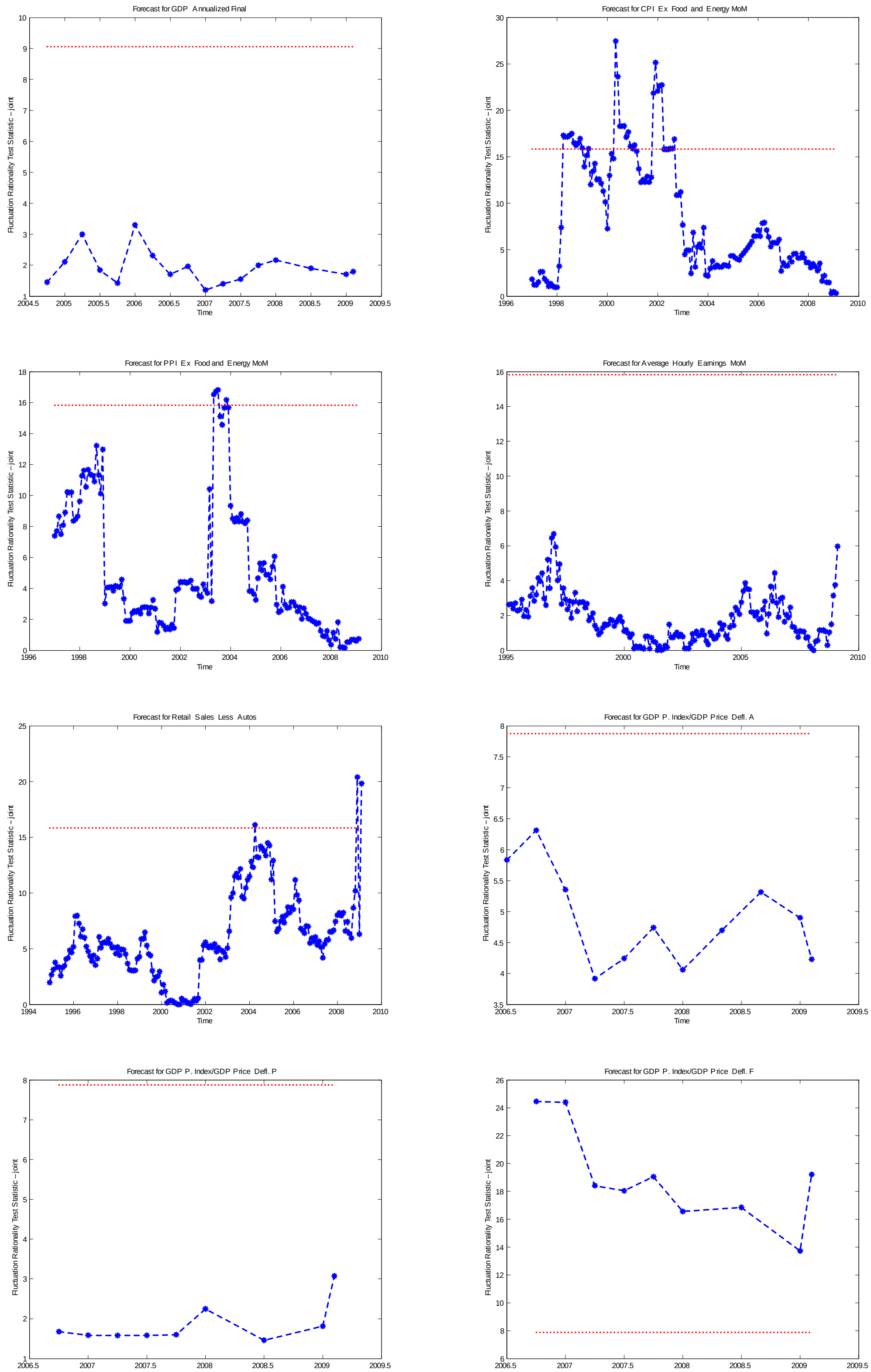

Forecast for Monthly Budget Statement
Fluctuation Rationality Test Statistic – joint
Time


Forecast for Consumer Credit
Fluctuation Rationality Test Statistic – joint
Time


Forecast for Initial Jobless Claims
Fluctuation Rationality Test Statistic – joint
Time


Forecast for GDP Annualized Advanced
Fluctuation Rationality Test Statistic – joint
Time


Forecast for GDP Annualized Preliminary
Fluctuation Rationality Test Statistic – joint
Time




Note. The same as for Figures 2,3, and 4.